

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problems Mailbox.**

THIS PAGE BLANK (USPTO)



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 108626

TO: Nita M Minnifield
Location: cm1/8a07/8e12
Art Unit: 1645
Tuesday, November 25, 2003
Case Serial Number: 09/298523

From: Paul Schulwitz
Location: Biotech-Chem Library
CM1-6B06
Phone: 305-1954

paul.schulwitz@uspto.gov

Search Notes

Examiner Minnifield,

See attached results.

If you have any questions about this search feel free to contact me at any time.

Thank you for using STIC search services!

Paul Schulwitz
Technical Information Specialist
STIC Biotech/Chem Library
(703)305-1954

THIS PAGE BLANK (USPTO)

108626

From: Chan, Christina
Sent: Wednesday, November 19, 2003 11:51 AM
To: Minnifield, Nita; STIC-Biotech/ChemLib
Subject: RE: rush sequence search request

RECEIVED

NOV 19 2003

STIC-BIOTECH/ChemLib
(STIC)

Please rush. Thanks Chris

Chris Chan

TC 1600 New Hire Training Coordinator and SPE 1644
308-3973
CM-1, 9B19

-----Original Message-----

From: Minnifield, Nita
Sent: Wednesday, November 19, 2003 11:30 AM
To: Chan, Christina
Subject: rush sequence search request

Christina, please approve, 2 month amdt.

STIC
09/298523

Please do a commercial and interference sequence search on SEQ ID
NO: 13 and aa 263-442 of SEQ ID NO: 13.

Please provide a paper copy of the results.

Thanks,

Nita M. Minnifield
Art Unit 1645
Office CM1-8A07
Mailbox CM1-8E12
703-305-3394

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 11/21
Date Completed: 11/25
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:30:37 ; Search time 21.4457 Seconds
(without alignments)
1008.166 Million cell updates/sec

Title: US-09-298-523C-13

Perfect score: 2537

Sequence: 1 MFAKSERKHYHSIRKFSVG.....NRLTQQPPKTEKPAQSTP 511

Scoring table:

BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 328717 seqs, 42310658 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Issued Patents AA: *
1: /cgn2_6/ptodata/1/aa/5A.COMB.pep: *
2: /cgn2_6/ptodata/1/aa/5B.COMB.pep: *
3: /cgn2_6/ptodata/1/aa/6A.COMB.pep: *
4: /cgn2_6/ptodata/1/aa/6B.COMB.pep: *
5: /cgn2_6/ptodata/1/aa/PTUTS.COMB.pep: *
6: /cgn2_6/ptodata/1/aa/backfile1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2194.5	86.5	446	4	US-09-286-981B-9 Sequence 9, Appli
2	2177.5	85.8	446	4	US-09-286-981B-6 Sequence 6, Appli
3	1728	68.1	564	3	US-09-308-022-6 Sequence 6, Appli
4	1512.5	59.6	406	4	US-09-286-981B-18 Sequence 18, Appli
5	1506	59.4	631	3	US-08-847-065-25 Sequence 25, Appli
6	1455.5	57.4	426	4	US-09-286-981B-12 Sequence 12, Appli
7	1447.5	57.1	425	4	US-09-286-981B-13 Sequence 13, Appli
8	1446.5	57.0	424	4	US-09-286-981B-14 Sequence 14, Appli
9	1440.5	56.8	428	4	US-09-286-981B-7 Sequence 7, Appli
10	1432	56.4	425	4	US-09-286-981B-11 Sequence 11, Appli
11	1412.5	55.7	412	4	US-09-286-981B-17 Sequence 17, Appli
12	1410.5	55.6	419	4	US-09-286-981B-15 Sequence 15, Appli
13	1405.5	55.4	414	4	US-09-286-981B-16 Sequence 16, Appli
14	1401.5	55.2	414	4	US-09-286-981B-10 Sequence 10, Appli
15	1395	55.0	431	4	US-09-286-981B-3 Sequence 3, Appli
16	1373	54.1	453	3	US-08-961-083-38 Sequence 38, Appli
17	1373	54.1	453	4	US-09-536-784-38 Sequence 38, Appli
18	1291	50.9	413	4	US-09-286-981B-5 Sequence 5, Appli
19	1207	47.6	864	4	US-08-714-741-40 Sequence 40, Appli
20	1157	45.6	588	4	US-08-714-741-42 Sequence 42, Appli
21	904	35.6	605	4	US-08-714-741-46 Sequence 46, Appli
22	856.5	33.8	8991	4	US-08-714-741-32 Sequence 32, Appli
23	764	30.1	1231	4	US-08-714-741-41 Sequence 41, Appli
24	752	29.6	623	4	US-08-714-741-47 Sequence 47, Appli
25	714.5	28.2	251	4	US-09-286-981B-4 Sequence 4, Appli
26	636	25.1	142	3	US-08-847-065-21 Sequence 21, Appli
27	560	22.1	219	4	US-09-286-981B-8 Sequence 8, Appli

28	542	21.4	114	4	US-09-286-981B-19 Sequence 19, Appli
29	525.5	20.7	110	3	US-08-961-083-102 Sequence 102, App
30	525.5	20.7	110	4	US-09-536-784-102 Sequence 102, App
31	499.5	19.7	695	1	US-08-127-499A-23 Sequence 23, Appli
32	499.5	19.7	695	1	US-08-482-847-23 Sequence 23, Appli
33	491	19.4	619	1	US-08-465-746-2 Sequence 2, Appli
34	491	19.4	619	1	US-08-214-164-2 Sequence 2, Appli
35	491	19.4	619	2	US-08-467-852A-3 Sequence 3, Appli
36	491	19.4	619	2	US-08-246-636-2 Sequence 2, Appli
37	491	19.4	619	2	US-08-247-491A-3 Sequence 3, Appli
38	491	19.4	619	2	US-08-319-795-2 Sequence 2, Appli
39	491	19.4	619	2	US-08-468-985-2 Sequence 2, Appli
40	491	19.4	619	3	US-08-312-949-2 Sequence 2, Appli
41	491	19.4	648	1	US-08-072-070-2 Sequence 2, Appli
42	491	19.4	648	1	US-08-469-434-2 Sequence 2, Appli
43	491	19.4	648	1	US-08-214-222-2 Sequence 2, Appli
44	491	19.4	648	2	US-08-467-852A-2 Sequence 2, Appli
45	491	19.4	648	2	US-08-468-718-2 Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-286-981B-9

Sequence 9, Application US/09286981B
Patent No. 6503511

GENERAL INFORMATION:

APPLICANT: Wizenmann, Theresa M.

APPLICANT: Koenig, Scott

TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines

Query Match 86.5%; Score 2194.5; DB 4; Length 446;

Best Local Similarity 99.8%; Pred. No. 7.3e-146;

Matches 445; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY	37	TENEGATQVPTSSNRNANESQAEQGEOPKUDSRDKARKVEEYVKKIVGESYAKTKKR	96
DB	1	TENEGATQVPTSSNRNANESQAEQGEOPKUDSRDKARKVEEYVKKIVGESYAKTKKR	60
QY	97	HTTVALVNLNINIKNELKIVESTSESQOLIMESREKVNBAVAFKPKSSSSSSD	156
DB	61	HTTVALVNLNINIKNELKIVESTSESQOLIMESREKVNBAVAFKPKSSSSSSD	120
QY	157	SSTPEASDTAKNNKPTPEGEKVAEAKKVEEA-KKAKDQKEEDRRNYPITYTTLLEI	215
DB	121	SSTPEASDTAKNNKPTPEGEKVAEAKKVEEA-KKAKDQKEEDRRNYPITYTTLLEI	180
QY	216	AESDVEYKAELELVYKKNANPEPDQIKQAEAEVSKQAEARLKKIKTDREAEAEAK	275
DB	181	AESDVEYKAELELVYKKNANPEPDQIKQAEAEVSKQAEARLKKIKTDREAEAEAK	240
QY	276	RRADAEQKPKGRKRGVGEIATPDKKENDAKSSDSSVGEETLPSPSLKPEKVAEAE	335
DB	241	RRADAEQKPKGRKRGVGEIATPDKKENDAKSSDSSVGEETLPSPSLKPEKVAEAE	300
QY	336	KKVEAKKKAEDQKEEDRRNYPITYTTLLEIAESDVEYKAELELVKKEAEPRNEEK	395
DB	301	KKVEAKKKAEDQKEEDRRNYPITYTTLLEIAESDVEYKAELELVKKEAEPRNEEK	360

QY 396 VKQAKAEVSKKAEATRLKIKTDKKAEEBAKKAEEADYKKEKPAQOPAPAPKAEK 455
DB 361 VKQAKAEVSKKAEATRLKIKTDKKAEEBAKKAEEADYKKEKPAQOPAPAPKAEK 420
QY 456 PAPAPKPPNPAEQPKAEKPADQQAEE 481
DB 421 PAPAPKPPNPAEQPKAEKPADQQAEE 446

RESULT 2
US-09-286-981B-6
; Sequence 6, Application US/09286981B
; Patent No. 6503511
; GENERAL INFORMATION:
; APPLICANT: Witzemann, Theresa M.
; APPLICANT: Koenig, Scott
; APPLICANT: Johnson, Leslie S
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
; FILE REFERENCE: 469201-396
; CURRENT APPLICATION NUMBER: US/09/286,981B
; CURRENT FILING DATE: 1999-04-06
; PRIOR APPLICATION NUMBER: US 60/085,743
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-286-981B-6

Query Match 85.8%; Score 2177.5; DB 4; Length 446;
Best Local Similarity 98.9%; Pred. No. 1.1e-144;
Matches 441; Conservative 1; Mismatches 3; Indels 1; Gaps 1;
QY 37 TENEGATVPTSSNRANESQAEQEGQPKKLDSEBDKAEVEEYKIKVGSYAKSTYKR 96
DB 1 TENEGATVPTSSNRANESQAEQEGQPKKLDSEBDKAEVEEYKIKVGSYAKSTYKR 60
QY 97 HTITVALVNEINNIKNIEYLINKIVESTSSQIILMMESRSKYDEAVSKFEKSSSSSSSD 156
DB 61 HTITVALVNEINNIKNIEYLINKIVESTSSQIILMMESRSKYDEAVSKFEKSSSSSSSD 120
QY 157 SSTPEASDTAKPNKPTPEGEKVAEAKKVEEA-KKADQKEPRRNPPTTYKTLLEI 215
DB 121 SSTPEASDTAKPNKPTPEGEKVAEAKKVEEA-KKADQKEPRRNPPTTYKTLLEI 180
QY 216 AESDVEVKALELVKVANEPDEQIKQAEAEVSKQAEATRLKIKTDREAEAEBAK 275
DB 181 AESDVEVKALELVKVANEPDEQIKQAEAEVSKQAEATRLKIKTDREAEAEBAK 240
QY 276 RRAADKEQKPKRGKRGVPGELATPDKKENDAKSSDSSVGEETLPSPSLKPEKKVAEAE 335
DB 241 RRAADKEQKPKRGKRGVPGELATPDKKENDAKSSDSSVGEETLPSPSLKPEKKVAEAE 300
QY 336 KKVBEAKKKAADQKEDRRNPPTNYKTLLEIAESDVEVKALELVKAEKPRNEEK 395
DB 301 KKVBEAKKKAADQKEDRRNPPTNYKTLLEIAESDVEVKALELVKAEKPRNEEK 360
QY 396 VKQAKAEVSKKAEATRLKIKTDKKAEEBAKKAEEADYKKEKPAQOPAPAPKAEK 455
DB 361 VKQAKAEVSKKAEATRLKIKTDKKAEEBAKKAEEADYKKEKPAQOPAPAPKAEK 420
QY 456 PAPAPKPPNPAEQPKAEKPADQQAEE 481
DB 421 PAPAPKPPNPAEQPKAEKPADQQAEE 446

RESULT 3
US-09-308-022-6
; Sequence 6, Application US/09308022
; Patent No. 6291654
; GENERAL INFORMATION:

APPLICANT: REGENTS OF THE UNIVERSITY OF MINNESOTA, ET AL.
TITLE OF INVENTION: C3 BINDING PROTEIN OF STREPTOCOCCUS
TITLE OF INVENTION: PNEUMONIAE
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: MUEITING, RAASCH & GEBHARDT, P.A.
STREET: 119 No. 6291654th Fourth Street, Suite 203
CITY: Minneapolis
STATE: Minnesota
COUNTRY: USA
ZIP: 55401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/308,022
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/20586
FILING DATE: 12-NOV-1997
APPLICATION NUMBER: 60/029,444
FILING DATE: 12-NOV-1996
APPLICATION NUMBER: 60/038,086
FILING DATE: 18-FEB-1997
APPLICATION NUMBER: 60/059,368
FILING DATE: 19-SEP-1997
APPLICATION NUMBER: 60/062,473
FILING DATE: 16-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: MUEITING, Ann M.
REGISTRATION NUMBER: 33,977
REFERENCE/DOCKET NUMBER: 110.00430101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-305-1217
TELEFAX: 612-305-1228
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 564 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-308-022-6

Query Match 68.1%; Score 1728; DB 3; Length 564;
Best Local Similarity 70.5%; Pred. No. 3.8e-113;
Matches 370; Conservative 50; Mismatches 47; Indels 58; Gaps 10;
QY 1 MFASKSRKRVYSIRKRSVGVASVVAASLVGVSVAHTENEGATVPTSSNRANESQAEQ 59
DB 1 MFASKSRKRVYSIRKRSVGVASVVAASLVGVSVAHTENEGATVPTSSNRANESQAEQ 54
QY 60 GEQPKKLDSEBDKAEK-VEEYVYKIKVGSYAKSTKRHTITVALVNEINNIKNIEYLINK- 117
DB 55 -----TEHRKKAQVNBVEIEKMLRE--IQDRKHTQVNLNKLKSLIKTKYREL 104
QY 118 -IVESTSSQIILMMESRSKYDEAVSKFEKSSSSSSDSTKPEASDTAKPNKPTPEG 176
DB 105 NVLEBKSKDELPR--SIRKAKLDAFRKFKDT-----LKPQ 138
QY 177 EKVAEAKKVEEA-KKADQKEEDRRNPPTTYKTLLEIAESDVEVKALELVKVKAN 235
DB 139 EKVAEAKKVEEA-KKADQKEEDRRNPPTNYKTLLEIAEPDVVKAEALVKEBAK 198
QY 236 EPRDEQIKQAEAEVSKQAEATRLKIKTDREAEAEBAKRAKAK-----EQGP 286
DB 199 ESRNEGITKQAEAEVSKQAEATRLKIKTDKKAEEBAKKAADYKKEANVATSDQGP 258
QY 287 KGRKRGVPGELATPDKKENDAKSSDSSVGEETLPSPSLKPEKKVAEAEKVEBAKKAEE 346

Db 259 KGRKRGVPGELATPDCKENDAKSSDSVGEETLPSLSLKGKVAEAKVEAEKAK 318
Qy 347 DQKEDRRNYPNTYKTTLELEIAESDVYKAELELVYVEAEPRNEEVKQAKAEVSK 406
Db 319 DQKEDRRNYPNTYKTTLELEIAESDVYKAELELVYVEAEPRDEEKIKQAKAEVSK 378
Qy 407 KAEATRLERIKITDRKKAEEBAERKAAEEDVKEKEPAEOPAPAPAKAEKAPAPENPA 466
Db 379 KAEATRLERIKITDRKKAEEBAERKAAEEDVKEKEPAEOPAPAPAKAEKAPAPENPA 436
Qy 467 EOPKAEKADQAEEDYARRSEEVNRLTQOOPPTKPAOPSTP 511
Db 437 EOPKAEKADQAEEDYARRSEEVNRLTQOOPPTKPAOPSTP 481

RESULT 4

US-09-286-981B-18
; Sequence 18, Application US/09286981B
; Patent No. 6503511
; GENERAL INFORMATION:
; APPLICANT: Witzmann, Theresa M.
; APPLICANT: Koenig, Scott
; APPLICANT: Johnson, Leslie S
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
; FILE REFERENCE: 469201-396
; CURRENT APPLICATION NUMBER: US/09/286,981B
; PRIOR FILING DATE: 1999-04-06
; PRIOR APPLICATION NUMBER: US 60/085,743
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 406
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-286-981B-18

Query Match 59.6%; Score 1512.5; DB 4; Length 406;
Best Local Similarity 73.0%; Pred. No. 2.9e-98;
Matches 327; Conservative 28; Mismatches 46; Indels 47; Gaps 8;

Qy 37 TENEGATQVPTSNRANESQAEQEGOPKDLSEDRDKAEVEEYVKIAGESVAKSTYKR 96
Db 1 TENEGTQAPTSSNRGNSQA-----EHMKAKAKQVDEVIEKML-----QDLRRK 44
Qy 97 HTTVALVNLNINIKETLKIIVSTESQQLIMMESRSKVDAVSKFEKSSSSSSSD 156
Db 45 HTQVNGLLTGLKATKTELRGLSVSKKSTAE-LPSEIKETLTAAPKQFKK----- 94
Qy 157 SSTPEASDTAKPKPTPEPGEKVAEAKKVEEA-KKAKDQKEEDRRNYPNTYKTTLELEI 215
Db 95 -----DLTKPK-----KVAEAEKVAEAKKADQKEEDRRNYPNTYKTTLELEI 140
Qy 216 AESDVYKAELELVYKANEPRDEQIKQAEAVESKQAEATRLKIKITDRSEAEAEK 275
Db 141 AESDVYKAELELVYKANEPRDEQIKQAEAVESKQAEATRLKIKITDRSEAEAEK 200
Qy 276 RRADAKQG---KPKGRKRGVPGELATPDCKENDAKSSDSVGEETLPSLSLKPKKY 331
Db 201 RRVVAKQDESSKRSKRVKRGDLGEQATPDCKENDAKSSDSVGEETLPSLSLKPKKY 260
Qy 332 AEAETKVEAKKKKADQKEEDRRNYPNTYKTTLELEIAESDVYKAELELVYVEAEPR 391
Db 261 AEAETKVEAKKKKADQKEEDRRNYPNTYKTTLELEIAESDVYKAELELVYVEAEPR 320
Qy 392 NEEKVQAKAEVSKKAEATRLKIKITDRKKAEEBAERKAAEEDVKEKEPAEOPAPAP 451
Db 321 NEEKVQAKAEVSKKAEATRLKIKITDRKKAEEBAERKAAEEDVKEKEPAEOPAPAP 380
Qy 452 KAEKPAAPKPEPNPAEOPKAEKPADQQA 479
Db 381 QPEK--PAPKPEPNPAEOPKAEKPADQQA 406

RESULT 5

US-08-847-065-25
; Sequence 25, Application US/08847065
; Patent No. 6245335
; GENERAL INFORMATION:
; APPLICANT: Masure, H. Robert
; APPLICANT: Rosencow, Carsten I.
; APPLICANT: Tuomanen, Blaine
; APPLICANT: Witzmann, Theresa M.
; TITLE OF INVENTION: CHOLINE BINDING PROTEINS FOR
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; STREET: Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/847,065
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-158
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 631 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: N-terminal
US-08-847-065-25

Query Match 59.4%; Score 1506; DB 3; Length 631;
Best Local Similarity 73.5%; Pred. No. 1.4e-97;
Matches 316; Conservative 39; Mismatches 35; Indels 40; Gaps 6;

Qy 94 KKRHTTVALVNLNINIKETLKIIVSTESQQLIMMESRSKVDAVSKFEKSSSS 151
Db 9 RRRHTQVNLNINILSAIKTYLRLVNLVLEKSKDELPE---SELKATLDAFAEFKDT-- 63
Qy 152 SSSSDSYTPASDTAKPKPTPEPGEKVAEAKKVEEA-KKAKDQKEEDRRNYPNTYKTT 210
Db 64 -----LKPGEKVAEAKKVEEAKKADQKEEDRRNYPNTYKTT 102
Qy 211 LELEIAESDVYKAELELVYKANEPRDEQIKQAEAVESKQAEATRLKIKITDRSEAE 270
Db 103 LELEIAESDVYKAELELVYKANEPRDEQIKQAEAVESKQAEATRLKIKITDRSEAE 162
Qy 271 EBEAKRRADAK-----EQGPKGRKRGVPGELATPDCKENDAKSSDSVGEETLP 321
Db 163 EBEAKRRADAKLEAVATSDQCKPKGRKRGVPGELATPDCKENDAKSSDSVGEETLP 222
Qy 322 SPSLKPEKKVAEAEKKEKKAEDQKEEDRRNYPNTYKTTLELEIAESDVYKAELE 381
Db 223 SSSLKSGKVAEAEKKEKKAEDQKEEDRRNYPNTYKTTLELEIAESDVYKAELE 282
Qy 382 LVVEAEKPEPRNEEKVQAKAEVSKKAEATRLKIKITDRKKAEEBAERKAAEEDVKEKEP 441

```
Db 283 LVKEAEKPEHDEEKKIKQAKAVESKKAATLENIKTDRKKAEEAKKAEEKXVEK 342
Qy 442 AEOPAPAPAKPEKPAKPEKPAKPEKPAKPEKPAKPEKPAKPEKPAKPEKPAK 501
Db 343 AEOPAPAPATOPPEK-PAPKEPEKPAKPEKPAKPEKPAKPEKPAKPEKPAKPEK 400
Qy 502 TEKPAPSTP 511
Db 401 TEKPAPSTP 410
```

RESULT 6
US-09-286-981B-12

```
; Sequence 12, Application US/09286981B
; Patent No. 6503511
; GENERAL INFORMATION:
; APPLICANT: Wizemann, Theresa M.
; APPLICANT: Koenig, Scott
; APPLICANT: Johnson, Leslie S
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
; FILE REFERENCE: 469201-396
; CURRENT APPLICATION NUMBER: US/09/286,981B
; CURRENT FILING DATE: 1999-04-06
; PRIOR APPLICATION NUMBER: US 60/085,743
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 12
; LENGTH: 426
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-286-981B-12
```

Query Match 57.4%; Score 1455.5; DB 4; Length 426;
Best Local Similarity 68.4%; Pred. No. 3e-94;
Matches 318; Conservative 41; Mismatches 47; Indels 59; Gaps 9;

```
Qy 37 TENEGATQVPTSSNRANESQAEQEPKLDSEKDKARKEVEEYVKIIVGESYAKSTKR 96
Db 1 TEKEVTTQVATSSSKAKNSQ-----TEHMKAKQVDEYIKKKI-----QLDRK 44
Qy 97 HTITVALVNEIINKIENYLNKIVESTESQQLIMESRSKVDKAVSKFEKSSSSSSD 156
Db 45 HTQVGLITKLGIVITETLHGLSVSKKSEAE-LPSEIKALDAFQPKDT----- 96
Qy 157 SSTPEASDTAKPMPTEPGEKVAEAKKVEEA-KKADQKEEDRRNYPTTYKTLELEI 215
Db 97 -----LPTEPGKVAEAEKKEVBAKKAEDQEKDRLNYPNTYKTLELEI 142
Qy 216 AESDVEYKAELELVKQANEPEDQKIKQAEAEVESQAEATRLKIKITDRKEAEK 275
Db 143 AESDVEYKAELELVKQANEPEDQKIKQAEAEVESQAEATRLKIKITDRKEAEK 201
Qy 276 RRADAK-----EQGPKGRARGVPGELATPDKKENDAKSSDSSVGEETLPSPLK 326
Db 202 RRADAKQAEANVATSEODKSKRAKREVLGELATPDKKENDAKSSDSSVGEETLPSPLK 261
Qy 327 PEKVAAEAEKKEVBAKKAEDQKEEDRRNYPTTYKTLELEIAESDVEYKAELELVKEE 386
Db 262 PEKVAAEAEKKEVBAKKAEDQKEEDRRNYPTTYKTLELEIAESDVEYKAELELVKEE 321
Qy 387 AKERNNEKIKQVAKAVESKKAATRLKIKITDRKKA-EEBAKRAAEEDVYKKEPAOP 445
Db 322 AKERNNEKIKQVAKAVESKKAATRLKIKITDRKKA-EEBAKRAAEEDVYKKEPAOP 381
Qy 446 QPAPAPAEK-----PAPAPKPEKPEKPAKPEKPAKPEKPAKPEKPAKPEK 481
Db 382 QPAPAPPEKPEKPEKPEKPAKPEKPAKPEKPAKPEKPAKPEKPAKPEKPAK 426
```

RESULT 7
US-09-286-981B-13

```
; Sequence 13, Application US/09286981B
; Patent No. 6503511
; GENERAL INFORMATION:
; APPLICANT: Wizemann, Theresa M.
; APPLICANT: Koenig, Scott
; APPLICANT: Johnson, Leslie S
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
; FILE REFERENCE: 469201-396
; CURRENT APPLICATION NUMBER: US/09/286,981B
; CURRENT FILING DATE: 1999-04-06
; PRIOR APPLICATION NUMBER: US 60/085,743
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 13
; LENGTH: 425
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-286-981B-13
```

Query Match 57.1%; Score 1447.5; DB 4; Length 425;
Best Local Similarity 68.3%; Pred. No. 1.1e-93;
Matches 317; Conservative 41; Mismatches 47; Indels 59; Gaps 9;

```
Qy 37 TENEGATQVPTSSNRANESQAEQEPKLDSEKDKARKEVEEYVKIIVGESYAKSTKR 96
Db 1 TEKEVTTQVATSSSKAKNSQ-----TEHMKAKQVDEYIKKKI-----QLDRK 44
Qy 97 HTITVALVNEIINKIENYLNKIVESTESQQLIMESRSKVDKAVSKFEKSSSSSSD 156
Db 45 HTQVGLITKLGIVITETLHGLSVSKKSEAE-LPSEIKALDAFQPKDT----- 96
Qy 157 SSTPEASDTAKPMPTEPGEKVAEAKKVEEA-KKADQKEEDRRNYPTTYKTLELEI 215
Db 97 -----LPTEPGKVAEAEKKEVBAKKAEDQEKDRLNYPNTYKTLELEI 142
Qy 216 AESDVEYKAELELVKQANEPEDQKIKQAEAEVESQAEATRLKIKITDRKEAEK 275
Db 143 AESDVEYKAELELVKQANEPEDQKIKQAEAEVESQAEATRLKIKITDRKEAEK 201
Qy 276 RRADAK-----EQGPKGRARGVPGELATPDKKENDAKSSDSSVGEETLPSPLK 326
Db 202 RRADAKQAEANVATSEODKSKRAKREVLGELATPDKKENDAKSSDSSVGEETLPSPLK 261
Qy 327 PEKVAAEAEKKEVBAKKAEDQKEEDRRNYPTTYKTLELEIAESDVEYKAELELVKEE 386
Db 262 PEKVAAEAEKKEVBAKKAEDQKEEDRRNYPTTYKTLELEIAESDVEYKAELELVKEE 321
Qy 387 AKERNNEKIKQVAKAVESKKAATRLKIKITDRKKA-EEBAKRAAEEDVYKKEPAOP 445
Db 322 AKERNNEKIKQVAKAVESKKAATRLKIKITDRKKA-EEBAKRAAEEDVYKKEPAOP 381
Qy 446 QPAPAPAEK-----PAPAPKPEKPEKPAKPEKPAKPEKPAKPEKPAKPEK 480
Db 382 QPAPAPPEKPEKPEKPEKPAKPEKPAKPEKPAKPEKPAKPEKPAKPEKPAK 425
```

RESULT 8
US-09-286-981B-14

```
; Sequence 14, Application US/09286981B
; Patent No. 6503511
; GENERAL INFORMATION:
; APPLICANT: Wizemann, Theresa M.
; APPLICANT: Koenig, Scott
; APPLICANT: Johnson, Leslie S
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
; FILE REFERENCE: 469201-396
; CURRENT APPLICATION NUMBER: US/09/286,981B
; CURRENT FILING DATE: 1999-04-06
; PRIOR APPLICATION NUMBER: US 60/085,743
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patentin Ver. 2.1
```

SEQ ID NO 14
LENGTH: 424
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-286-981B-14

Query Match 57.0%; Score 1446.5; DB 4; Length 424;
Best Local Similarity 68.5%; Pred. No. 1.2e-93;
Matches 317; Conservative 40; Mismatches 47; Indels 59; Gaps 9;

```
QY 37 TENEGATQVPTSSNRANESQAEOGQPKLDSERDARKVEEYVYKIVGESYAKSTKR 96
DB 1 TEKEVTTQVATSSNRANESQA-----TEHKKAKQVDEYIKKTL-----QIDRRK 44
QY 97 HTITVALVNLNNIKNEYLNKIVESTSESQOLIMMESRSKYDEAVSKPEKSSSSSSSD 156
DB 45 HTQVGLTKLGVITTEYHGLSVSKKSEAE-LPSEITAKIDAAFEQPKDT----- 96
QY 157 SSTKEASDTAKPNKPTERGKVAEAKKVEEA-KKADQKEEDRRNYPTIYTKLELEI 215
DB 97 -----LPTERGKVAEAKKVEEA-KKADQKEEDRRNYPTIYTKLELEI 142
QY 216 AESDVEVKKAELELVKVAKEPRDQKIKQAEAVESKQAEATRLKIKTDREAEAEAK 275
DB 143 AESDVEVKKAELELVKVAKEPRDQKIKQAEAVESKQAEATRLKIKTDREAEAEAK 201
QY 276 RRADAK-----EQQKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPSPSLK 326
DB 202 RRADAKLEQANVAISEQKSRARRAEVLGELATPDKKENDAKSSDSSVGEETLPSPSLK 261
QY 327 PEKVAEAEKVEAEKKAEDQKEEDRRNYPTIYTKLELEIAESDVEVKKAELELVKEE 386
DB 262 PEKVAEAEKVEAEKKAEDQKEEDRRNYPTIYTKLELEIAESDVEVKKAELELVKEE 321
QY 387 AKERNNEKIKQVAKAEVSKKAEATRLKIKTDREAEAEAKKAEEDYKKEKPAEQ 445
DB 322 AKERNNEKIKQVAKAEVSKKAEATRLKIKTDREAEAEAKKAEEDYKKEKPAEQ 381
QY 446 QAPAPKAEK-----PAPAPKPNPAOPKAEKPADQAA 479
DB 382 QAPAPKAEKPTPEENPAPAPKPNPAOPKAEKPADQAA 424
```

RESULT 9

US-09-286-981B-7
Sequence 7, Application US/09286981B
Patent No. 6503511
GENERAL INFORMATION:
APPLICANT: Witzmann, Theresa M.
APPLICANT: Koenig, Scott
APPLICANT: Johnson, Leslie S
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
FILE REFERENCE: 469201-396
CURRENT FILING DATE: 1999-04-06
PRIOR APPLICATION NUMBER: US 60/085,743
PRIOR FILING DATE: 1998-05-15
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 7
LENGTH: 428
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-286-981B-7

Query Match 56.8%; Score 1440.5; DB 4; Length 428;
Best Local Similarity 69.7%; Pred. No. 3.3e-93;
Matches 311; Conservative 35; Mismatches 53; Indels 47; Gaps 7;
QY 60 GEQPKLDSERDARK--EVEEYVYKIVGESYAKSTKRHTITVALVNLNNIKNEYLN 116
DB 6 GNNSTVTVSSGQDISKVADEVESHOSILKDVNKLKKVOHTQVADPNKLSIKIKTYLY 65

```
QY 117 K-IVESTSESQOLIMMESRSKYDEAVSKPEKSSSSSSDSSSTKPEASDTAKPNKPT 174
DB 66 ELNVLEEKSEAEELNSKTKETKEBELTAPEQPKDTLS-----TE 104
QY 175 PEKVAEAKKVEEA-KKAKQKEEDRRNYPTIYTKLELEIAESDVEVKKAELELVK 233
DB 105 PEKVAEAKKVEEA-KKAKQKEEDRRNYPTIYTKLELEIAESDVEVKKAELELVK 164
QY 234 ANEPRDEKIKQAEAVESKQAEATRLKIKTDREAE-----EAKRRADAK 282
DB 165 ANEPRDEKIKQAEAVESKQAEATRLKIKTDREAE-----EAKRRADAK 224
QY 283 QGKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPSPSLPEKVAEAEKVEEAK 342
DB 225 E-FKKTKRGVLGEPAITPDKKENDAKSSDSSVGEETLPSPSLPEKVAEAEKVEEAK 282
QY 343 KKAEDQKEEDRRNYPTIYTKLELEIAESDVEVKKAELELVKEAEKPRNEKVAQAAE 402
DB 283 KKAEDQKEEDRRNYPTIYTKLELEIAESDVEVKKAELELVKEAEKPRNEKVAQAAE 342
QY 403 VESKAEATRLKIKTDREAEAEAKKAEEDYKKEKPAEQOPAPAPKAEK 455
DB 343 VESKAEATRLKIKTDREAEAEAKKAEEDYKKEKPAEQOPAPAPKAEK 402
QY 456 PAPAPKPNPAOPKAEKPADQAAE 481
DB 403 PAPAPKPNPAOPKAEKPADQAAE 428
```

RESULT 10

US-09-286-981B-11
Sequence 11, Application US/09286981B
Patent No. 6503511
GENERAL INFORMATION:
APPLICANT: Witzmann, Theresa M.
APPLICANT: Koenig, Scott
APPLICANT: Johnson, Leslie S
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
FILE REFERENCE: 469201-396
CURRENT FILING DATE: 1999-04-06
PRIOR APPLICATION NUMBER: US 60/085,743
PRIOR FILING DATE: 1998-05-15
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 11
LENGTH: 425
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-286-981B-11

Query Match 56.4%; Score 1432; DB 4; Length 425;
Best Local Similarity 66.6%; Pred. No. 1.3e-92;
Matches 311; Conservative 44; Mismatches 48; Indels 64; Gaps 9;

```
QY 37 TENEGATQVPTSSNRANESQAEOGQPKLDSERDARKVEEYVYKIVGESYAKSTKR 96
DB 1 TEKEVTTQVATSSNRANESQA-----KAAEQFDEYIKTMI-----QIDRRK 44
QY 97 HTITVALVNLNNIKNEYLNK-IVESTSESQOLIMMESRSKYDEAVSKPEKSSSSSS 154
DB 45 HTQVGLTKLGSITKTYLKLKLVLEKSKAEELP--SETKKEIDAAFEQPKDTNRTTK 101
QY 155 SDSSTKPEASDTAKPNKPTERGKVAEAKKVEEA-KKADQKEEDRRNYPTIYTKLEL 213
DB 102 T-----VAEAEKVEEA-KKAKQKEEDRRNYPTIYTKLEL 138
QY 214 EIASDVEVKKAELELVKVAKEPRDQKIKQAEAVESKQAEATRLKIKTDREAEAE 273
DB 139 EIASDVEVKKAELELVKVAKEPRDQKIKQAEAVESKQAEATRLKIKTDREAEAE 198
QY 274 AKRRADAK-----EQGKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLP 321
```

Db 199 AKRBAEAKLKEAVEKNVATSEODKPKGRKRGVGEQATPPKENDAKSSDSSVGEALP 258
Qy 322 SPSLPEKKVAEAEKKVAEAKKAEEDKEDRRNPTTYKTLELAEISDVYKKALE 381
Db 259 SPSLPEKKVAEAEKKVAEAKKAEEDKEDRRNPTTYKTLELAEISDVYKKALE 318
Qy 382 LVKEAEKPRBEKKVAEAKKAEATRLKIKITDRKAAEEBKRAAEEDKVEKXP 441
Db 319 LVKEAEKSRBEKKVAEAKKAEATRLKIKITDRKAAEEBKRAAEEDKVEKXP 378
Qy 442 AEQOPAPAPKAEK-----PAPAKPENPAEQPKAEKPADQOAE 481
Db 379 AEQOPAPAPKAEKPEEPENPAAPKPEKPAEQPKAEKTDQOAE 425
RESULT 11
US-09-286-981B-17
; Sequence 17, Application US/09286981B
; Patent No. 6503511
; GENERAL INFORMATION:
; APPLICANT: Witzemann, Theresa M.
; APPLICANT: Koenig, Scott
; APPLICANT: Johnson, Leslie S
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
; FILE REFERENCE: 469201-396
; CURRENT APPLICATION NUMBER: US/09/286,981B
; PRIOR FILING DATE: 1999-04-06
; PRIOR APPLICATION NUMBER: US 60/085,743
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 17
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-286-981B-17
Query Match 55.7%; Score 1412.5; DB 4; Length 412;
Best Local Similarity 70.8%; Pred. No. 2.9e-91;
Matches 303; Conservative 41; Mismatches 51; Indels 33; Gaps 6;
Qy 63 PKLDSERDKARK--EVEEYVKIAGESYAKSTKRTITVALVNEINNIKNETLANK-- 117
Db 9 PTYSSGGDISKXADYKSHLEKLSIQNLDSKIKITVNLINKQDILKRTLYLVLN 68
Qy 118 IVESITSESOQLIMMESSKVDKAVSKREKSSSSSSSSSSSTKPEASDTAKPNKTEPGE 177
Db 69 VLEDKSKAELP---SKIKAEILDAAFEQPKDT-----LPTBPGK 104
Qy 178 KVAEAKKKVVEEA--KKAQKQKEDRRNPTTYKTLELAEISDVYKKALELVYKANE 236
Db 105 KVAEAKKKVVEEA--KKAQKQKEDRRNPTTYKTLELAEISDVYKKALELVYKANE 164
Qy 237 PRDEOKIKQAAEVESKQAEATRLKIKITDRKAAEEBKRAAEEDKVEKXP 293
Db 165 SRNEGTINQAKAVESQAEATRLKIKITDRKAAEEBKRAAEEDKVEKXP 224
Qy 294 VPGELATPDKKENDAKSSDSSVGEETLPSPLPEKKVAEAEKKVAEAKKAEEDKVEKXP 353
Db 225 DFGEPATPDKKENDAKSSDSSVGEETLPSPLPEKKVAEAEKKVAEAKKAEEDKVEKXP 284
Qy 354 RNYPTNYKTLELAEISDVYKKALELVYKKAEEBKPRNEKKVQAKAEVSKKAEATRL 413
Db 285 RNYPTTYKTLELAEISDVYKKALELVYKKAEEBKPRNEKKVQAKAEVSKKAEATRL 344
Qy 414 EKITDRKAAEEBKRAAEEDKVEKXPAPQOPAPAPKAEKPAEPENPAEQPKAEK 473
Db 345 EKITDRKAAEEBKRAAEEDKVEKXPAPQOPAPAPKAEKPAEPENPAEQPKAEK 404
Qy 474 PADQOAE 481
Db 405 PADQOAE 412

RESULT 12
US-09-286-981B-15
; Sequence 15, Application US/09286981B
; Patent No. 6503511
; GENERAL INFORMATION:
; APPLICANT: Witzemann, Theresa M.
; APPLICANT: Koenig, Scott
; APPLICANT: Johnson, Leslie S
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
; FILE REFERENCE: 469201-396
; CURRENT APPLICATION NUMBER: US/09/286,981B
; PRIOR FILING DATE: 1999-04-06
; PRIOR APPLICATION NUMBER: US 60/085,743
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 15
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-286-981B-15
Query Match 55.6%; Score 1410.5; DB 4; Length 419;
Best Local Similarity 66.5%; Pred. No. 4e-91;
Matches 308; Conservative 43; Mismatches 49; Indels 63; Gaps 9;
Qy 37 TENEGATQVPTSSNRANESQAEQOPKLDSEDKARKEVEEYVKIAGESYAKSTKR 96
Db 1 TENERTQVPTSSNRGR-----PERKKAAEQFDEYINKMI-----QLDKRK 41
Qy 97 HTITVALVNEINNIKNETLAEISDVYKKALELVYKKAEEBKPRNEKKVQAKAEVSKKAEATRLKIKITDR 268
Db 42 HTQNLAEINQSLRKIKITLYLNGLKEK--SEALP---SKIKAEILDAAFEQPKDT----- 90
Qy 157 SSTPEASDTAKPNKTEPGEKVAEAKKVEEA-----KKAQKQKEDRRNPTTY 208
Db 91 -----LPTPEKKVAEAEKKVAEAKKVAEAKKAKKAEEDKEDRRNPTTY 136
Qy 209 KTLLELAEISDVYKKALELVYKKAEEBKPRNEKKVQAKAEVSKKAEATRLKIKITDR 268
Db 137 KTLLELAEISDVYKKALELVYKKAEEBKPRNEGTINQAKAVESKAEATRLKIKITDR 196
Qy 269 EA--EEBKRRADAEQ--KPKGRKGVGELATPDKKENDAKSSDSSVGEETLPS 324
Db 197 KAESEAKRRADAEQDESRRKSRKRGALGEQATPDKENDAKSSDSSVGEETLPS 256
Qy 325 LKPEKVAEAEKKVAEAKKAEEDKEDRRNPTTYKTLELAEISDVYKKALELVYK 384
Db 257 LKPKVAEAEKKVAEADKKAKAEEDKEDRRNPTTYKTLELAEISDVYKKALELVYK 316
Qy 385 EEAKEPRNEKKVQAKAEVSKKAEATRLKIKITDRKAAEEBKRAAEEDKVEKXP 444
Db 317 EEAESNEEKIKQAKAVESKKAATRLKIKITDRKAAEEBKRAAEEDKVEKXP 376
Qy 445 POPAPAPKAEK-----PAPKPEPAEQPKAEKPADQOAE 480
Db 377 POPAPAPKAEKPEEPENPAEPKPEPAEQPKAEKPADQOAE 419
RESULT 13
US-09-286-981B-16
; Sequence 16, Application US/09286981B
; Patent No. 6503511
; GENERAL INFORMATION:
; APPLICANT: Witzemann, Theresa M.
; APPLICANT: Koenig, Scott
; APPLICANT: Johnson, Leslie S
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
; FILE REFERENCE: 469201-396
; CURRENT APPLICATION NUMBER: US/09/286,981B
; PRIOR FILING DATE: 1999-04-06
; PRIOR APPLICATION NUMBER: US 60/085,743


```
/ PRIOR FILING DATE: 1998-05-15
/ NUMBER OF SEQ ID NOS: 38
/ SOFTWARE: Patentln Ver. 2.1
/ SEQ ID NO 16
/ LENGTH: 414
/ TYPE: PRT
/ ORGANISM: Streptococcus pneumoniae
US-09-286-981B-16

Query Match      55.4%; Score 1405.5; DB 4; Length 414;
Best Local Similarity 67.7%; Pred. No. 8.9e-91;
Matches 310; Conservative 45; Mismatches 46; Indels 57; Gaps 11;

Oy 37 TENEGATQVPTSSNRANESQAEOGQPKUDSERDKARKEVEEYVKTIAGESYAKSTKK 95
Db 1 TENEGSTQAATSSNNAK-----TEHKAAKQVVDYEITKMLRE--IQDLRR 44
Oy 96 RHITVALVNLNINKNEYLNK--IVESTSESQOILMMESRSKYDEAVSKFEKSSSSS 153
Db 45 KHTQVVALNINKLSAIKTKYLRNLVLEEKSKDELPE--SEIKAKLDAAFEFK----- 92
Oy 154 SSDSTKEASDTAKPNKPTPEGEKVAAKKVVEA-KKAKDQKEEDRRNPPTTYKTLE 212
Db 93 -----EKKDTLK-----PGEKVAAKKVVEAKKAAEDQKEEDRRNPPTNYKTLE 138
Oy 213 LEIAESDVEYKKAELVLYKVANEPDRDEOKIQAEEAVESKQAEATRLKKITDREBAE 272
Db 139 LEIAEFVYKVAEAELELVKEEAKESRNNGTIIKQAEKESKKAETRLLENITDKRKAAE 198
Oy 273 EAKRADAK-----EOKPKGRARGVPGELATPDKKENDAKSSDSSVGEETLPSP 323
Db 199 EAKRADAKLKEANVAATSDQKPKGRARGVPGELATPDKKENDAKSSDSSVGEETLPSS 258
Oy 324 SLKPEKVAEAEKVEAEKKAEDQKEEDRRNPPTNYKTLELEIAESDVEYKKAELV 383
Db 259 SLKSGKVAEAEKVEAEKKAEDQKEEDRRNPPTNYKTLELEIAESDVEYKKAELV 318
Oy 384 KEBAKEPNEEKVQAQAEVSKKAETRLKIKTRKKAEEBAKKAEDVKYKPAE 443
Db 319 KEBAKEPDEEKIKQAQAKVESKKAETRLLENITDKRKAEEBAKKAEDVKYKPAE 378
Oy 444 QOPAPAPKAEKPAAPKPPENPAEQPKAEKPADQAAE 481
Db 379 QOPAPATQPEK--PAPKEKPAEQPKAEKTDQAAE 414

RESULT 14
US-09-286-981B-10
/ Sequence 10, Application US/09286981B
/ Patent No. 6503511
/ GENERAL INFORMATION:
/ APPLICANT: Witzmann, Theresa M.
/ APPLICANT: Koenig, Scott
/ APPLICANT: Johnson, Leslie S
/ TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
/ FILE REFERENCE: 469201-136
/ CURRENT FILING DATE: 1999-04-06
/ PRIOR FILING DATE: 1998-05-15
/ NUMBER OF SEQ ID NOS: 38
/ SOFTWARE: Patentln Ver. 2.1
/ SEQ ID NO 10
/ LENGTH: 414
/ TYPE: PRT
/ ORGANISM: Streptococcus pneumoniae
US-09-286-981B-10

Query Match      55.2%; Score 1401.5; DB 4; Length 414;
Best Local Similarity 66.6%; Pred. No. 1.7e-90;
Matches 305; Conservative 48; Mismatches 48; Indels 57; Gaps 9;

Oy 37 TENEGATQVPTSSNRANESQAEOGQPKUDSERDKARKEVEEYVKTIAGESYAKSTKK 95
```

```
Db 1 TENEGSTQAATSSNNAK-----TEHKAAKQVVDYEITKMLRE--IQDLRR 44
Oy 96 RHITVALVNLNINKNEYLNK--IVESTSESQOILMMESRSKYDEAVSKFEKSSSSS 153
Db 45 KHTQVVALNINKLSAIKTKYLRNLVLEEKSKDELPE--SEIKAKLDAAFEFK----- 97
Oy 154 SSDSTKEASDTAKPNKPTPEGEKVAAKKVVEA-KKAKDQKEEDRRNPPTTYKTLE 212
Db 98 -----LKPGEKVAAKKVVEAKKAAEDQKEEDRRNPPTNYKTLE 138
Oy 213 LEIAESDVEYKKAELVLYKVANEPDRDEOKIQAEEAVESKQAEATRLKKITDREBAE 272
Db 139 LEIAEFVYKVAEAELELVKEEAKESRNNGTIIKQAEKESKKAETRLLENITDKRKAAE 198
Oy 273 EAKRADAK-----EOKPKGRARGVPGELATPDKKENDAKSSDSSVGEETLPSP 323
Db 199 EAKRADAKLKEANVAATSDQKPKGRARGVPGELATPDKKENDAKSSDSSVGEETLPSS 258
Oy 324 SLKPEKVAEAEKVEAEKKAEDQKEEDRRNPPTNYKTLELEIAESDVEYKKAELV 383
Db 259 SLKSGKVAEAEKVEAEKKAEDQKEEDRRNPPTNYKTLELEIAESDVEYKKAELV 318
Oy 384 KEBAKEPNEEKVQAQAEVSKKAETRLKIKTRKKAEEBAKKAEDVKYKPAE 443
Db 319 KEBAKEPDEEKIKQAQAKVESKKAETRLLENITDKRKAEEBAKKAEDVKYKPAE 378
Oy 444 QOPAPAPKAEKPAAPKPPENPAEQPKAEKPADQAAE 481
Db 379 QOPAPATQPEK--PAPKEKPAEQPKAEKTDQAAE 414

RESULT 15
US-09-286-981B-3
/ Sequence 3, Application US/09286981B
/ Patent No. 6503511
/ GENERAL INFORMATION:
/ APPLICANT: Witzmann, Theresa M.
/ APPLICANT: Koenig, Scott
/ APPLICANT: Johnson, Leslie S
/ TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
/ FILE REFERENCE: 469201-136
/ CURRENT FILING DATE: 1999-04-06
/ PRIOR FILING DATE: 1998-05-15
/ NUMBER OF SEQ ID NOS: 38
/ SOFTWARE: Patentln Ver. 2.1
/ SEQ ID NO 3
/ LENGTH: 431
/ TYPE: PRT
/ ORGANISM: Streptococcus pneumoniae
US-09-286-981B-3

Query Match      55.0%; Score 1395; DB 4; Length 431;
Best Local Similarity 64.0%; Pred. No. 5.1e-90;
Matches 304; Conservative 57; Mismatches 60; Indels 54; Gaps 9;

Oy 37 TENEGATQVPTSSNRANESQAEOGQPKUDSERDKARKEVEEYVKTIAGESYAKSTKK 96
Db 1 TEKEVTTTPVATSSSKANKSQ-----TEHKAAAEVDEYINMGI-----QDLKRX 44
Oy 97 HTITVALVNLNINKNEYLNKIVESSESQOILMMESRSKYDEAVSKFEKSSSSSSD 156
Db 45 HTQVVALNINKLSAIKTKYLRNLVLEEKSKDELTSKXKIDDAFQFNKDT----- 97
Oy 157 SSTKPEASDTAKPNKPTPEGEKVAAKKVVEA-KKAKDQKEEDRRNPPTTYKTLEI 215
Db 98 -----LKPGEKVAAKKVVEAKKAAEDQKEEDRRNPPTNYKTLEI 141
Oy 216 AESDVEYKKAELVLYKVANEPDRDEOKIQAEEAVESKQAEATRLKKITDREBAEBAK 275
Db 142 AESDVEYKKAELVLYVEBAKGSNEEKIKQAQAEVSKKAETRLKIKTRKKAEEBAK 201
```

Qy	276	RRADAKQ--GKPKGRARQVPGELATPDKENDAKSSDSSVGEETLPSISL-----	326
Db	202	RKAEEVEVKNNLKKRTRGAFGEPAIPDKENDAKSSDSSVVKKS-SKPIIKSEKVAE	260
Qy	327	PEKVAEAEKKEVEAKKAEDQKEEDRRNYPNTYKTLLEIAESDVEKKALELVKEE	386
Db	261	AEKVAEAEKKEVEAKKAEDQKEEDRRNYPNTYKTLLEIAESDVKKEAELELVKEE	320
Qy	387	AKPPNEEKVKQAKAEVSKAEATRLKIKITDRKKAEEAKRKAEEEDKVEKPAEQPQ	446
Db	321	AKPPNEEKIKQAKAEVSKAEATRLKIKITDRKKA-EBAKRVAEEDKVEKPAEQPQ	379
Qy	447	PAPAPKAEKPAAPKPEENPAEQPKAEKPADQQAEEYARSSSEYNRLLTQOQPK	501
Db	380	PAPAP--KPAAPQPEKPAEQPKAEKPADQQAEEYARSSSEYNRLLTAPAK	431

Search completed: November 21, 2003, 13:35:10
 Job time : 22.4457 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:30:37 ; Search time 7.55427 Seconds
(without alignments)
1008.166 Million cell updates/sec

Title: US-09-298-523C-13_COPY_263_442

Perfect score: 889

Sequence: 1 IKTDREAEAEAKRADAKE.....AEEAKKAEEDKVEKPA 180

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/aa/5A.COMB.pep:*
- 2: /cgn2_6/ptodata/1/aa/5B.COMB.pep:*
- 3: /cgn2_6/ptodata/1/aa/6A.COMB.pep:*
- 4: /cgn2_6/ptodata/1/aa/6B.COMB.pep:*
- 5: /cgn2_6/ptodata/1/aa/PTCUTS.COMB.pep:*
- 6: /cgn2_6/ptodata/1/aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	889	100.0	446	4	US-09-286-981B-9 Sequence 9, Appl
2	889	100.0	453	3	US-08-961-083-38 Sequence 38, Appl
3	889	100.0	453	4	US-09-536-784-38 Sequence 38, Appl
4	884	99.4	446	4	US-09-286-981B-6 Sequence 6, Appl
5	803.5	90.4	414	4	US-09-286-981B-16 Sequence 16, Appl
6	803.5	90.4	631	3	US-08-847-065-25 Sequence 25, Appl
7	801.5	90.2	564	3	US-09-308-022-6 Sequence 6, Appl
8	792	89.1	425	4	US-09-286-981B-18 Sequence 18, Appl
9	792	89.1	425	4	US-09-286-981B-11 Sequence 11, Appl
10	791.5	89.0	414	4	US-09-286-981B-10 Sequence 10, Appl
11	784	88.2	428	4	US-09-286-981B-7 Sequence 7, Appl
12	775.5	87.2	412	4	US-09-286-981B-17 Sequence 17, Appl
13	770.5	86.7	424	4	US-09-286-981B-14 Sequence 14, Appl
14	770.5	86.7	426	4	US-09-286-981B-12 Sequence 12, Appl
15	769.5	86.6	425	4	US-09-286-981B-13 Sequence 13, Appl
16	768	86.4	419	4	US-09-286-981B-15 Sequence 15, Appl
17	717.5	80.7	413	4	US-09-286-981B-5 Sequence 5, Appl
18	684.5	77.0	864	4	US-08-714-741-40 Sequence 40, Appl
19	676.5	76.1	431	4	US-09-286-981B-3 Sequence 3, Appl
20	669	75.3	588	4	US-08-714-741-42 Sequence 42, Appl
21	557.5	62.7	219	4	US-09-286-981B-8 Sequence 8, Appl
22	542	61.0	114	4	US-09-286-981B-19 Sequence 19, Appl
23	521	58.6	142	3	US-08-847-065-21 Sequence 21, Appl
24	517	58.2	251	4	US-09-286-981B-4 Sequence 4, Appl
25	487	54.8	103	4	US-09-286-981B-1 Sequence 1, Appl
26	461.5	51.9	605	4	US-08-714-741-46 Sequence 46, Appl
27	451	50.7	103	4	US-09-286-981B-38 Sequence 38, Appl

28	447.5	50.3	1231	4	US-08-714-741-41 Sequence 41, Appl
29	394.5	44.4	623	4	US-08-714-741-47 Sequence 47, Appl
30	375.5	42.2	8991	4	US-08-714-741-32 Sequence 32, Appl
31	306	34.4	110	3	US-08-961-083-102 Sequence 102, Appl
32	306	34.4	110	4	US-09-536-784-102 Sequence 102, Appl
33	237.5	26.7	55	4	US-09-286-981B-37 Sequence 37, Appl
34	182.5	20.5	1507	3	US-08-929-329-5 Sequence 5, Appl
35	174	19.6	468	4	US-09-328-352-6321 Sequence 6321, Appl
36	159.5	17.9	611	4	US-09-216-393B-81 Sequence 81, Appl
37	158.5	17.8	288	3	US-08-312-949-4 Sequence 4, Appl
38	158.5	17.8	288	3	US-08-446-201-4 Sequence 4, Appl
39	158.5	17.8	619	1	US-08-465-746-2 Sequence 2, Appl
40	158.5	17.8	619	1	US-08-214-164-2 Sequence 2, Appl
41	158.5	17.8	619	2	US-08-467-852A-3 Sequence 3, Appl
42	158.5	17.8	619	2	US-08-246-636-2 Sequence 2, Appl
43	158.5	17.8	619	2	US-08-247-691A-3 Sequence 3, Appl
44	158.5	17.8	619	2	US-08-319-795-2 Sequence 2, Appl
45	158.5	17.8	619	2	US-08-468-985-2 Sequence 2, Appl

ALIGNMENTS

```
RESULT 1
US-09-286-981B-9
Sequence 9, Application US/09286981B
Patent No. 6503511
GENERAL INFORMATION:
APPLICANT: Wizemann, Theresa M.
APPLICANT: Koenig, Scott
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
FILE REFERENCE: 469201-396
CURRENT APPLICATION NUMBER: US/09/286,981B
PRIOR FILING DATE: 1999-04-06
PRIOR APPLICATION NUMBER: US 60/085,743
NUMBER OF SEQ ID NOS: 38
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 9
LENGTH: 446
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-286-981B-9

Query Match 100.0%; Score 889; DB 4; Length 446;
Best Local Similarity 100.0%; Pred. No. 1.2e-65;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IKTDREAEAEAKRADAKEOGKPKGRAGVGEIATDPDKENDAKSSSVGEITLPS 60
228 IKTDREAEAEAKRADAKEOGKPKGRAGVGEIATDPDKENDAKSSSVGEITLPS 287
Db 228 IKTDREAEAEAKRADAKEOGKPKGRAGVGEIATDPDKENDAKSSSVGEITLPS 287
QY 61 PSLKPKKVAEAKKVEAKKKAEDOKEDRRVYPNTYKTLLEIAESDVEYKAELEL 120
288 PSLKPKKVAEAKKVEAKKKAEDOKEDRRVYPNTYKTLLEIAESDVEYKAELEL 347
Db 288 PSLKPKKVAEAKKVEAKKKAEDOKEDRRVYPNTYKTLLEIAESDVEYKAELEL 347
QY 121 VKEAEAPNEEKKVOKAEVSKKAETRLKIKTRKKAEEAKRKAEDVKVEKPA 180
348 VKEAEAPNEEKKVOKAEVSKKAETRLKIKTRKKAEEAKRKAEDVKVEKPA 407
Db 348 VKEAEAPNEEKKVOKAEVSKKAETRLKIKTRKKAEEAKRKAEDVKVEKPA 407

RESULT 2
US-08-961-083-38
Sequence 38, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSER: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
```

CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961.083
FILING DATE:
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 453 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-38

Query Match 100.0%; Score 889; DB 3; Length 453;
Best Local Similarity 100.0%; Pred. No. 1.3e-65;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPPDKKENDAKSSDSVGEETLPS 60
25 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPPDKKENDAKSSDSVGEETLPS 84
61 PSLEPEKVAEAEKKEVEAKKKAEDQKEEDRRNYPNTYKTLLEIAESDVEVKAELEL 120
85 PSLEPEKVAEAEKKEVEAKKKAEDQKEEDRRNYPNTYKTLLEIAESDVEVKAELEL 144
121 VKEAKEPRNEEKVKQAKAEVESKKAETRLKIKTRDKKAAEEAKRAAEEDVKKEKPA 180
145 VKEAKEPRNEEKVKQAKAEVESKKAETRLKIKTRDKKAAEEAKRAAEEDVKKEKPA 204

RESULT 3
US-09-536-784-38
Sequence 38, Application US/09536784
Patent No. 6573082
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/536,784
FILING DATE: 30-Oct-1997
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961.083
FILING DATE: OCT-30-1997
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB340P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 453 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 38:
US-09-536-784-38

Query Match 100.0%; Score 889; DB 4; Length 453;
Best Local Similarity 100.0%; Pred. No. 1.3e-65;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPPDKKENDAKSSDSVGEETLPS 60
25 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPPDKKENDAKSSDSVGEETLPS 84
61 PSLEPEKVAEAEKKEVEAKKKAEDQKEEDRRNYPNTYKTLLEIAESDVEVKAELEL 120
85 PSLEPEKVAEAEKKEVEAKKKAEDQKEEDRRNYPNTYKTLLEIAESDVEVKAELEL 144
121 VKEAKEPRNEEKVKQAKAEVESKKAETRLKIKTRDKKAAEEAKRAAEEDVKKEKPA 180
145 VKEAKEPRNEEKVKQAKAEVESKKAETRLKIKTRDKKAAEEAKRAAEEDVKKEKPA 204

RESULT 4
US-09-286-981B-6
Sequence 6, Application US/09286981B
Patent No. 6503511
GENERAL INFORMATION:
APPLICANT: Witzmann, Theresa M.
APPLICANT: Koenig, Scott
APPLICANT: Johnson, Leslie S
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
FILE REFERENCE: 469201-396
CURRENT APPLICATION NUMBER: US/09/286,981B
CURRENT FILING DATE: 1999-04-06
PRIOR APPLICATION NUMBER: US 60/085,743
PRIOR FILING DATE: 1998-05-15
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 6
LENGTH: 446
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-286-981B-6

Query Match 99.4%; Score 884; DB 4; Length 446;
Best Local Similarity 99.4%; Pred. No. 3.2e-65;
Matches 179; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPPDKKENDAKSSDSVGEETLPS 60
228 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPPDKKENDAKSSDSVGEETLPS 287
61 PSLEPEKVAEAEKKEVEAKKKAEDQKEEDRRNYPNTYKTLLEIAESDVEVKAELEL 120
288 PSLEPEKVAEAEKKEVEAKKKAEDQKEEDRRNYPNTYKTLLEIAESDVEVKAELEL 347
121 VKEAKEPRNEEKVKQAKAEVESKKAETRLKIKTRDKKAAEEAKRAAEEDVKKEKPA 180

Db 348 VKEAKEPRNEKVKQAKAEVSKAEATRLKIKTDKRAKAEBAKRAAEEDKVKRPA 407

RESULT 5
US-09-286-981B-16
Sequence 16, Application US/09286981B
Patent No. 6503511
GENERAL INFORMATION:
APPLICANT: Witzmann, Theresa M.
APPLICANT: Koenig, Scott
APPLICANT: Johnson, Leslie S
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
FILE REFERENCE: 469201-396
CURRENT APPLICATION NUMBER: US/09/286,981B
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: US 60/085,743
NUMBER OF SEQ ID NOS: 38
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 16
LENGTH: 414
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-286-981B-16

Query Match 90.4%; Score 803.5; DB 4; Length 414;
Best Local Similarity 86.8%; Pred. No. 1.2e-58;
Matches 164; Conservative 12; Mismatches 4; Indels 9; Gaps 1;

Qy 1 IKTDREAEAEBAKRRADAK-----EOKPKGRAGVPGELATPPDKKENDAKSSDS 51
Db 189 IKTRKKAEEBAKRRADAKLKEANVATSDQKPKGRAGVPGELATPPDKKENDAKSSDS 248

Qy 52 SVGEETLPSPSLKPEKKVAEAEKKVKEAKKADQKEDRRNYPNTYKTLLEIAESDV 111
Db 249 SVGEETLPSSSLKSGKVAEAEKKVKEAKKADQKEDRRNYPNTYKTLLEIAESDV 308

Qy 112 EVKKAELVKEBAKEPRNEKVKQAKAEVSKAEATRLKIKTDKRAKAEBAKRAAE 171
Db 309 KVKAELVKEBAKEPRDEKIKQAKAKVESKAEATRLLENIKTDKRAKAEBAKRAAE 368

Qy 172 EDKVKRPA 180
Db 369 EDKVKRPA 377

RESULT 6
US-08-847-065-25
Sequence 25, Application US/08847065
Patent No. 6245335
GENERAL INFORMATION:
APPLICANT: Masure, H. Robert
APPLICANT: Rosenow, Carsten I.
APPLICANT: Tuomanen, Elaine
APPLICANT: Witzmann, Theresa M.
TITLE OF INVENTION: CHOLINE BINDING PROTEINS FOR
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: David A. Jackson, Esq.
STREET: 411 Hackensack Ave, Continental Plaza, 4th
STREET: Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/847,065

FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-158
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 631 amino acids
TYPE: amino acid
STRANDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
FRAGMENT TYPE: N-terminal
US-08-847-065-25

Query Match 90.4%; Score 803.5; DB 3; Length 631;
Best Local Similarity 86.8%; Pred. No. 2e-58;
Matches 164; Conservative 12; Mismatches 4; Indels 9; Gaps 1;

Qy 1 IKTDREAEAEBAKRRADAK-----EOKPKGRAGVPGELATPPDKKENDAKSSDS 51
Db 155 IKTRKKAEEBAKRRADAKLKEANVATSDQKPKGRAGVPGELATPPDKKENDAKSSDS 214

Qy 52 SVGEETLPSPSLKPEKKVAEAEKKVKEAKKADQKEDRRNYPNTYKTLLEIAESDV 111
Db 215 SVGEETLPSSSLKSGKVAEAEKKVKEAKKADQKEDRRNYPNTYKTLLEIAESDV 274

Qy 112 EVKKAELVKEBAKEPRNEKVKQAKAEVSKAEATRLKIKTDKRAKAEBAKRAAE 171
Db 275 KVKAELVKEBAKEPRDEKIKQAKAKVESKAEATRLLENIKTDKRAKAEBAKRAAE 334

Qy 172 EDKVKRPA 180
Db 335 EDKVKRPA 343

RESULT 7
US-09-308-022-6
Sequence 6, Application US/09308022
Patent No. 6291654
GENERAL INFORMATION:
APPLICANT: REGENTS OF THE UNIVERSITY OF MINNESOTA, ET AL.
TITLE OF INVENTION: C3 BINDING PROTEIN OF STREPTOCOCCUS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: MUETING, RAASCH & GEBHARDT, P.A.
STREET: 119 No. 6291654th Fourth Street, Suite 203
CITY: Minneapolis
STATE: Minnesota
COUNTRY: USA
ZIP: 55401

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/308,022
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/20586
FILING DATE: 12-NOV-1997
APPLICATION NUMBER: 60/029,444
FILING DATE: 12-NOV-1996
APPLICATION NUMBER: 60/038,086
FILING DATE: 18-FEB-1997

APPLICATION NUMBER: 60/059,368
FILING DATE: 19-SEP-1997
APPLICATION NUMBER: 60/062,473
FILING DATE: 16-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: MUEYING, Ann M.
REGISTRATION NUMBER: 33,977
REFERENCE/DOCKET NUMBER: 110.00430101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-305-1217
TELEFAX: 612-305-1228
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 564 amino acids
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-308-022-6

Query Match 90.2%; Score 801.5; DB 3; Length 564;
Best Local Similarity 86.2%; Pred. No. 2.5e-58;
Matches 163; Conservative 13; Mismatches 4; Indels 9; Gaps 1;

QY 1 IKTDRBEAEAEKRRADAK-----EOKPKGRARGVPGELATPDKKENDAKSSDS 51
DB 226 IKTDRBEAEAEKRRADAKLKEANVATSDQKPKGRARGVPGELATPDKKENDAKSSDS 285
QY 52 SVGEETLPSPLKPEKKVAEAEKKVEAKKAEDQKEEDRRNYPNTYKTLLEIAESDV 111
DB 286 SVGEETLPSPLSKSGKVAEAEKKVEAEKKAEDQKEEDRRNYPNTYKTLLEIAESDV 345
QY 112 EVKKAELVVEAEKPPNEKVKQAQAEVSKQAELRLKIKTDRKAAEEAKRAE 171
DB 346 KYKEAELEIVVEAEKPPDEEKIKQAQAKVESKKAELRLKIKTDRKAAEEAKRAE 405
QY 172 EDKVEKKA 180
DB 406 EDKVEKKA 414

RESULT 8
US-09-286-981B-18
Sequence 18, Application US/09286981B
Patent No. 6503511
GENERAL INFORMATION:
APPLICANT: Wizemann, Theresa M.
APPLICANT: Koenig, Scott
APPLICANT: Johnson, Leslie S
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
FILE REFERENCE: 469201-396
CURRENT APPLICATION NUMBER: US/09/286,981B
CURRENT FILING DATE: 1999-04-06
PRIOR APPLICATION NUMBER: US 60/085,743
PRIOR FILING DATE: 1998-05-15
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 18
LENGTH: 406
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-286-981B-18

Query Match 89.1%; Score 792; DB 4; Length 406;
Best Local Similarity 90.2%; Pred. No. 1e-57;
Matches 166; Conservative 3; Mismatches 11; Indels 4; Gaps 1;

QY 1 IKTDRBEAEAEKRRADAKG---KPKGRARGVPGELATPDKKENDAKSSDSVCEE 56
DB 188 IKTDRBEAEAEKRRADAKGDESSKRRKRVKRGDLQOATPDKKENDAKSSDSVCEE 247
QY 57 TLPSPSLKPEKKVAEAEKKVEAKKAEDQKEEDRRNYPNTYKTLLEIAESDVAKKA 116

DB 248 TLPSPSLKPGKVAEAEKKVEAEADKKAQAKQKEEDRRNYPNTYKTLLEIAESDVAKKA 307
QY 117 ELEIVKEAEKPPNEKVKQAQAEVSKQAELRLKIKTDRKAAEEAKRAEEDKVK 176
DB 308 ELEIVKEAEKPPNEKVKQAQAEVSKQAELRLKIKTDRKAAEEAKRAEEDKVK 367
QY 177 EKPA 180
DB 368 EKPA 371

RESULT 9
US-09-286-981B-11
Sequence 11, Application US/09286981B
Patent No. 6503511
GENERAL INFORMATION:
APPLICANT: Wizemann, Theresa M.
APPLICANT: Koenig, Scott
APPLICANT: Johnson, Leslie S
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
FILE REFERENCE: 469201-396
CURRENT APPLICATION NUMBER: US/09/286,981B
CURRENT FILING DATE: 1999-04-06
PRIOR APPLICATION NUMBER: US 60/085,743
PRIOR FILING DATE: 1998-05-15
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 11
LENGTH: 425
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-286-981B-11

Query Match 89.1%; Score 792; DB 4; Length 425;
Best Local Similarity 85.4%; Pred. No. 1.1e-57;
Matches 164; Conservative 8; Mismatches 8; Indels 12; Gaps 1;

QY 1 IKTDRBEAEAEKRRADAK-----EOKPKGRARGVPGELATPDKKENDAKS 48
DB 188 IKTDRBEAEAEKRRADAKLKEAVKNAVTSQKPKGRARGVPGELATPDKKENDAKS 247
QY 49 SDSSVGEETLPSPLKPEKKVAEAEKKVEAKKAEDQKEEDRRNYPNTYKTLLEIAE 108
DB 248 SDSSVGEETLPSPLKPEKKVAEAEKKVEAKKAEDQKEEDRRNYPNTYKTLLEIAE 307
QY 109 SDVEKKAELVVEAEKPPNEKVKQAQAEVSKQAELRLKIKTDRKAAEEAKRK 168
DB 308 SDVAKSELEIVVEAEKPPNEKVKQAQAEVSKQAELRLKIKTDRKAAEEAKRK 367
QY 169 AAEDKVEKKA 180
DB 368 AAEDKVEKKA 379

RESULT 10
US-09-286-981B-10
Sequence 10, Application US/09286981B
Patent No. 6503511
GENERAL INFORMATION:
APPLICANT: Wizemann, Theresa M.
APPLICANT: Koenig, Scott
APPLICANT: Johnson, Leslie S
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
FILE REFERENCE: 469201-396
CURRENT APPLICATION NUMBER: US/09/286,981B
CURRENT FILING DATE: 1999-04-06
PRIOR APPLICATION NUMBER: US 60/085,743
PRIOR FILING DATE: 1998-05-15
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 10
LENGTH: 414
TYPE: PRT

ORGANISM: Streptococcus pneumoniae
US-09-286-981B-10

Query Match 89.0%; Score 791.5; DB 4; Length 414;
Best Local Similarity 85.7%; Pred. No. 1.2e-57;
Matches 162; Conservative 12; Mismatches 6; Indels 9; Gaps 1;

QY 1 IKTDREAEAEAKRRADAK-----EQQKPKGRAGVPGELATPDKKENDAKSSDS 51
DB 189 IKTDREAEAEAKRRADAK-----EQQKPKGRAGVPGELATPDKKENDAKSSDS 248
QY 52 SVGEETLPSPLKPEKKVAEAEKKVAAEKKAEDQKEDRRNYPNTYKTLLEIASDSV 111
DB 249 SVGEETLPSPLKPEKKVAEAEKKVAAEKKAEDQKEDRRNYPNTYKTLLEIASDSV 308
QY 112 EVKKAELVYKEAEKPRNEKVKQAAYEVESKKAETRLLEKIKTDKKAEEBAKRAAE 171
DB 309 KVKAELELVYKEAEKPRNEKVKQAAYEVESKKAETRLLEKIKTDKKAEEBAKRAAE 368
QY 172 EDKVKKEPA 180
DB 369 EDKVKKEPA 377

RESULT 11
US-09-286-981B-7
Sequence 7, Application US/09286981B
Patent No. 6503511

GENERAL INFORMATION:
APPLICANT: Wizemann, Theresa M.
APPLICANT: Koenig, Scott
APPLICANT: Johnson, Leslie S
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
FILE REFERENCE: 469201-396
CURRENT APPLICATION NUMBER: US/09/286,981B
CURRENT FILING DATE: 1999-04-06
PRIOR APPLICATION NUMBER: US 60/085,743
PRIOR FILING DATE: 1998-05-15
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 7
LENGTH: 428
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-286-981B-7

Query Match 88.2%; Score 784; DB 4; Length 428;
Best Local Similarity 90.6%; Pred. No. 5e-57;
Matches 163; Conservative 4; Mismatches 7; Indels 6; Gaps 1;

QY 1 IKTDREAEAEAKRRADAK-----EQQKPKGRAGVPGELATPDKKENDAKSSDSVGEETLPS 60
DB 209 IKTDREAEAEAKRRADAK-----EQQKPKGRAGVPGELATPDKKENDAKSSDSVGEETLPS 262
QY 61 PSLKPEKKVAEAEKKVAAEKKAEDQKEDRRNYPNTYKTLLEIASDSVVKKAEL 120
DB 263 PSLKPEKKVAEAEKKVAAEKKAEDQKEDRRNYPNTYKTLLEIASDSVVKKAEL 322
QY 121 VKEAEKPRNEKVKQAAYEVESKKAETRLLEKIKTDKKAEEBAKRAAEEDKVKKEPA 180
DB 323 VKEAEKPRNEKVKQAAYEVESKKAETRLLEKIKTDKKAEEBAKRAAEEDKVKKEPA 382

RESULT 12
US-09-286-981B-17
Sequence 17, Application US/09286981B
Patent No. 6503511
GENERAL INFORMATION:
APPLICANT: Wizemann, Theresa M.
APPLICANT: Koenig, Scott
APPLICANT: Johnson, Leslie S
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
FILE REFERENCE: 469201-396

CURRENT APPLICATION NUMBER: US/09/286,981B
CURRENT FILING DATE: 1999-04-06
PRIOR APPLICATION NUMBER: US 60/085,743
PRIOR FILING DATE: 1998-05-15
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 17
LENGTH: 412
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-286-981B-17

Query Match 87.2%; Score 775.5; DB 4; Length 412;
Best Local Similarity 89.6%; Pred. No. 2.4e-56;
Matches 164; Conservative 4; Mismatches 12; Indels 3; Gaps 1;

QY 1 IKTDREAEAEAKRRADAK-----EQQKPKGRAGVPGELATPDKKENDAKSSDSVGEET 57
DB 191 IKTDREAEAEAKRRADAK-----EQQKPKGRAGVPGELATPDKKENDAKSSDSVGEET 250
QY 58 LPSPLKPEKKVAEAEKKVAAEKKAEDQKEDRRNYPNTYKTLLEIASDSVVKKAEL 117
DB 251 LPSPLKPEKKVAEAEKKVAAEKKAEDQKEDRRNYPNTYKTLLEIASDSVVKKAEL 310
QY 118 LELVKEAEKPRNEKVKQAAYEVESKKAETRLLEKIKTDKKAEEBAKRAAEEDKVKKE 177
DB 311 LELVKEAEKPRNEKVKQAAYEVESKKAETRLLEKIKTDKKAEEBAKRAAEEDKVKKE 370
QY 178 KPA 180
DB 371 KPA 373

RESULT 13
US-09-286-981B-14
Sequence 14, Application US/09286981B
Patent No. 6503511
GENERAL INFORMATION:
APPLICANT: Wizemann, Theresa M.
APPLICANT: Koenig, Scott
APPLICANT: Johnson, Leslie S
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
FILE REFERENCE: 469201-396
CURRENT APPLICATION NUMBER: US/09/286,981B
CURRENT FILING DATE: 1999-04-06
PRIOR APPLICATION NUMBER: US 60/085,743
PRIOR FILING DATE: 1998-05-15
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 14
LENGTH: 424
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-286-981B-14

Query Match 86.7%; Score 770.5; DB 4; Length 424;
Best Local Similarity 87.4%; Pred. No. 6.4e-56;
Matches 166; Conservative 4; Mismatches 9; Indels 11; Gaps 3;

QY 1 IKTDREAEAEAKRRADAK-----EQQKPKGRAGVPGELATPDKKENDAKSSDS 51
DB 190 IKTDREAEAEAKRRADAK-----EQQKPKGRAGVPGELATPDKKENDAKSSDS 248
QY 52 SVGEETLPSPLKPEKKVAEAEKKVAAEKKAEDQKEDRRNYPNTYKTLLEIASDSV 111
DB 249 SVGEETLPSPLKPEKKVAEAEKKVAAEKKAEDQKEDRRNYPNTYKTLLEIASDSV 308
QY 112 EVKKAELVYKEAEKPRNEKVKQAAYEVESKKAETRLLEKIKTDKKAEEBAKRAAE 170
DB 309 EVKKAELVYKEAEKPRNEKVKQAAYEVESKKAETRLLEKIKTDKKAEEBAKRAAE 368
QY 171 EEDKVKKEPA 180

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:33:52 ; Search time 39.1939 Seconds
(without alignments)
2380.162 Million cell updates/sec

Title: US-09-298-523C-13
Perfect score: 2537
Sequence: 1 MFASKSRKHYHSIRKFSVG.....NRLTQQQPKTEKPAQSTP 511

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues
Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	* Query Match	Length	DB ID	Description
1	2537	100.0	511	11	US-09-298-523B-13
2	2526	99.6	513	11	US-09-298-523B-12
3	2516	99.2	693	11	US-09-769-787-185
4	2350.5	92.6	655	11	US-09-056-019-2
5	2252.5	88.8	460	11	US-09-056-019-38
6	2235.5	88.1	459	11	US-09-056-019-39
7	2194.5	86.5	446	12	US-10-254-995-9
8	2177.5	85.8	446	12	US-10-254-995-6
9	2105	83.0	451	11	US-09-298-523B-67
10	2101.5	82.8	428	11	US-09-056-019-24
11	1979.5	78.0	406	11	US-09-056-019-1
12	1761.5	69.4	487	11	US-09-298-523B-9
13	1745.5	68.8	487	11	US-09-298-523B-66
14	1718	67.7	581	11	US-09-298-523B-56
15	1715	67.6	483	11	US-09-298-523B-10

16	1700	67.0	481	11	US-09-298-523B-6	Sequence 6, App1
17	1679.5	66.2	496	11	US-09-298-523B-4	Sequence 4, App1
18	1569.5	61.9	663	11	US-09-298-523B-58	Sequence 58, App1
19	1569.5	61.9	663	11	US-09-056-019-8	Sequence 8, App1
20	1569.5	61.9	663	12	US-09-969-748C-14	Sequence 14, App1
21	1569.5	61.9	663	12	US-09-949-039-104	Sequence 31, App1
22	1568.5	61.8	419	11	US-09-056-019-31	Sequence 31, App1
23	1559.5	61.5	478	11	US-09-298-523B-65	Sequence 65, App1
24	1534	60.5	437	11	US-09-298-523B-64	Sequence 64, App1
25	1526	60.1	444	11	US-09-298-523B-5	Sequence 5, App1
26	1512.5	59.6	406	12	US-10-254-995-18	Sequence 18, App1
27	1506	59.4	631	12	US-09-829-382-25	Sequence 25, App1
28	1504.5	59.3	439	11	US-09-056-019-28	Sequence 28, App1
29	1497.5	59.0	437	11	US-09-056-019-32	Sequence 32, App1
30	1496.5	59.0	437	11	US-09-056-019-29	Sequence 29, App1
31	1494	58.9	439	11	US-09-056-019-30	Sequence 30, App1
32	1466.5	57.8	439	11	US-09-056-019-37	Sequence 37, App1
33	1465.5	57.8	433	11	US-09-056-019-33	Sequence 33, App1
34	1455.5	57.4	426	12	US-10-254-995-12	Sequence 12, App1
35	1447.5	57.1	425	12	US-10-254-995-13	Sequence 13, App1
36	1446.5	57.0	424	12	US-10-254-995-14	Sequence 14, App1
37	1440.5	56.8	428	12	US-10-254-995-7	Sequence 7, App1
38	1436.5	56.6	425	12	US-09-056-019-36	Sequence 36, App1
39	1432	56.4	425	12	US-10-254-995-11	Sequence 11, App1
40	1412.5	55.7	412	12	US-10-254-995-17	Sequence 17, App1
41	1410.5	55.6	419	12	US-10-254-995-15	Sequence 15, App1
42	1408.5	55.5	413	11	US-09-056-019-35	Sequence 35, App1
43	1405.5	55.4	414	12	US-10-254-995-16	Sequence 16, App1
44	1401.5	55.2	414	12	US-10-254-995-10	Sequence 10, App1
45	1399	55.1	711	11	US-09-298-523B-3	Sequence 3, App1

ALIGNMENTS

RESULT 1
US-09-298-523B-13
; Sequence 13, Application US/09298523B
; Publication No. US20030059438A1
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OR INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; TITLE OR INVENTION: AND STRAINS THEREOF AND USES THEREFOR
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/298, 523B
; CURRENT FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-298-523B-13

Query Match 100.0%; Score 2537; DB 11; Length 511;
Best Local Similarity 100.0%; Pred. No. 3.9e-121;
Matches 511; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MFASKSRKHYHSIRKFSVGASVVAASLVWGSAVHAHTEGATOVPTSSNRANESQAEQG 60
DB 1 MFASKSRKHYHSIRKFSVGASVVAASLVWGSAVHAHTEGATOVPTSSNRANESQAEQG 60
QY EQPKKIDSEBDKARKEVEEYKKIVGSAKSTYKHTTITVALVNEIINIKNEYLNKIVE 120
DB EQPKKIDSEBDKARKEVEEYKKIVGSAKSTYKHTTITVALVNEIINIKNEYLNKIVE 120
QY STEESQLOILMMSRSKVDKVAASKFEKDSSSSSSTKPEASDPAKPKTEPGKVA 180
DB STEESQLOILMMSRSKVDKVAASKFEKDSSSSSSTKPEASDPAKPKTEPGKVA 180
QY 181 EAKKVEAKKAKADQEEEDRRNPITTYKTLEIAESDVEVKALELVKVAANPRDE 240
DB 181 EAKKVEAKKAKADQEEEDRRNPITTYKTLEIAESDVEVKALELVKVAANPRDE 240

```
Oy 241 QKIQAEAEVSKQAEATRLKKITDREAEAEAKRRADAEQCKPKRARKGVPGELAT 300
Db 241 QKIQAEAEVSKQAEATRLKKITDREAEAEAKRRADAEQCKPKRARKGVPGELAT 300
Oy 301 PDKKENDAKSSDSVGEETLPSPSLKPEKVAEAKVEAKKAEADQKEEDRRNYPTNT 360
Db 301 PDKKENDAKSSDSVGEETLPSPSLKPEKVAEAKVEAKKAEADQKEEDRRNYPTNT 360
Oy 361 YKTLLEIAESDVEVKAELELVKEAKEPRNEEKVKQAKAEVBSKKAATRLKIKITDR 420
Db 361 YKTLLEIAESDVEVKAELELVKEAKEPRNEEKVKQAKAEVBSKKAATRLKIKITDR 420
Oy 421 KKAEEAEAKRAAEEDVKVEKPAEQOPAPAPAKEPAPKPEBPAPQAEKPADQOAE 480
Db 421 KKAEEAEAKRAAEEDVKVEKPAEQOPAPAPAKEPAPKPEBPAPQAEKPADQOAE 480
Oy 481 EDVARSSEEEYNRLTQOQPPKTEKPAQPSSTP 511
Db 481 EDVARSSEEEYNRLTQOQPPKTEKPAQPSSTP 511
```

RESULT 2

```
US-09-298-523B-12
; Sequence 12, Application US/09298523B
; Publication No. US20030059438A1
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; FILE REFERENCE: 454312-3140
; CURRENT FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 513
; TYPE: PRN
; ORGANISM: Streptococcus pneumoniae
US-09-298-523B-12
```

```
Query Match 99.6%; Score 2526; DB 11; Length 513;
Best Local Similarity 99.6%; Pred. No. 1.4e-120;
Matches 509; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
Oy 1 MFASKSERKHYHSIRKFSVGVASVVVASLVMGSVVHATENEGATQVPTSSNRANESQAEQ 60
Db 1 MFASKSERKHYHSIRKFSVGVASVVVASLVMGSVVHATENEGATQVPTSSNRANESQAEQ 60
Oy 61 EOPKRLDSEBDKARKEVEEYVKKIYGESYAKSTKKRHTITVALVNLNINIKNEYLINKIV 120
Db 61 EOPKRLDSEBDKARKEVEEYVKKIYGESYAKSTKKRHTITVALVNLNINIKNEYLINKIV 120
Oy 121 STSESQLOILMMESRSKVDKAVSKPEKSSSSSSSSSTKPEASDTAKPNKPTPEGEKVA 180
Db 121 STSESQLOILMMESRSKVDKAVSKPEKSSSSSSSSSTKPEASDTAKPNKPTPEGEKVA 180
Oy 181 EAKKVEEAKKADQKEEDRRNYPTITYKTLLEIAESDVEVKAELELVKVKANEPDE 240
Db 181 EAKKVEEAKKADQKEEDRRNYPTITYKTLLEIAESDVEVKAELELVKVKANEPDE 240
Oy 241 QKIQAEAEVSKQAEATRLKKITDREAEAEAKRRADAEQCKPKRARKGVPGELAT 300
Db 241 QKIQAEAEVSKQAEATRLKKITDREAEAEAKRRADAEQCKPKRARKGVPGELAT 300
Oy 301 PDKKENDAKSSDSVGEETLPSPSLKPEKVAEAKVEAKKAEADQKEEDRRNYPTNT 360
Db 301 PDKKENDAKSSDSVGEETLPSPSLKPEKVAEAKVEAKKAEADQKEEDRRNYPTNT 360
Oy 361 YKTLLEIAESDVEVKAELELVKEAKEPRNEEKVKQAKAEVBSKKAATRLKIKITDR 420
Db 361 YKTLLEIAESDVEVKAELELVKEAKEPRNEEKVKQAKAEVBSKKAATRLKIKITDR 420
```

```
Oy 421 KKAEEAEAKRAAEEDVKVEKPAEQOPAPAPAKEPAPKPEBPAPQAEKPADQOAE 480
Db 421 KKAEEAEAKRAAEEDVKVEKPAEQOPAPAPAKEPAPKPEBPAPQAEKPADQOAE 480
Oy 481 EDVARSSEEEYNRLTQOQPPKTEKPAQPSSTP 511
Db 481 EDVARSSEEEYNRLTQOQPPKTEKPAQPSSTP 511
```

RESULT 3

```
US-09-769-787-185
; Sequence 185, Application US/09769787
; Publication No. US20030091577A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129WO
; CURRENT FILING DATE: 2001-01-26
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 185
; LENGTH: 693
; TYPE: PRN
; ORGANISM: Streptococcus pneumoniae
US-09-769-787-185
```

```
Query Match 99.2%; Score 2516; DB 11; Length 693;
Best Local Similarity 99.6%; Pred. No. 6.1e-120;
Matches 511; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
```

```
Oy 1 MFASKSERKHYHSIRKFSVGVASVVVASLVMGSVVHATENEGATQVPTSSNRANESQAEQ 59
Db 1 MFASKSERKHYHSIRKFSVGVASVVVASLVMGSVVHATENEGATQVPTSSNRANESQAEQ 60
Oy 60 GEOPKRLDSEBDKARKEVEEYVKKIYGESYAKSTKKRHTITVALVNLNINIKNEYLINKIV 119
Db 60 GEOPKRLDSEBDKARKEVEEYVKKIYGESYAKSTKKRHTITVALVNLNINIKNEYLINKIV 120
Oy 120 ESTSESQLOILMMESRSKVDKAVSKPEKSSSSSSSSSTKPEASDTAKPNKPTPEGEKV 179
Db 120 ESTSESQLOILMMESRSKVDKAVSKPEKSSSSSSSSSTKPEASDTAKPNKPTPEGEKV 180
Oy 180 AEAKKVEEAK-KKADQKEEDRRNYPTITYKTLLEIAESDVEVKAELELVKVKANEPDE 238
Db 180 AEAKKVEEAKKADQKEEDRRNYPTITYKTLLEIAESDVEVKAELELVKVKANEPDE 240
Oy 239 DEQKIQAEAEVSKQAEATRLKKITDREAEAEAKRRADAEQCKPKRARKGVPGEL 298
Db 241 DEQKIQAEAEVSKQAEATRLKKITDREAEAEAKRRADAEQCKPKRARKGVPGEL 300
Oy 299 ATPDKKENDAKSSDSVGEETLPSPSLKPEKVAEAKVEAKKAEADQKEEDRRNYPT 358
Db 301 ATPDKKENDAKSSDSVGEETLPSPSLKPEKVAEAKVEAKKAEADQKEEDRRNYPT 360
Oy 359 NTYKTLLEIAESDVEVKAELELVKEAKEPRNEEKVKQAKAEVBSKKAATRLKIKIT 418
Db 361 NTYKTLLEIAESDVEVKAELELVKEAKEPRNEEKVKQAKAEVBSKKAATRLKIKIT 420
Oy 419 DRKAEAEAKRAAEEDVKVEKPAEQOPAPAPAKEPAPKPEBPAPQAEKPADQOAE 478
Db 421 DRKAEAEAKRAAEEDVKVEKPAEQOPAPAPAKEPAPKPEBPAPQAEKPADQOAE 480
Oy 479 AEEDVARSSEEEYNRLTQOQPPKTEKPAQPSSTP 511
Db 481 AEEDVARSSEEEYNRLTQOQPPKTEKPAQPSSTP 513
```

```
RESULT 4
US-09-056-019-2
; Sequence 2, Application US/09056019A
; Publication No. US20030096950A1
; GENERAL INFORMATION:
; APPLICANT: Tuomanen, Elaine I
; APPLICANT: Wizemann, Theresa
; APPLICANT: Masure, H. R.
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Koenig, Scott
; TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL
; TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED
; FILE REFERENCE: 1340-1-017
; CURRENT FILING DATE: 1998-04-07
; CURRENT APPLICATION NUMBER: US/09/056,019A
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-056-019-2

Query Match      92.6%; Score 2350.5; DB 11; Length 655;
Best Local Similarity 99.8%; Pred. No. 14e-111;
Matches 474; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 38 ENEGATOVPTSSNRANESQAEQEQPKLDSERDKARKEVEYVKIVGESYAKSTKRRH 97
DB 1 ENEGATOVPTSSNRANESQAEQEQPKLDSERDKARKEVEYVKIVGESYAKSTKRRH 60
QY 98 TITVALVNELNINKIYKIVESTSESQLOIIMESRSKYDEAVSKVEKSDSSSSSDS 157
DB 61 TITVALVNELNINKIYKIVESTSESQLOIIMESRSKYDEAVSKVEKSDSSSSSDS 120
QY 158 STKPSASTAKPNKTEPGEKVAEAKKVEEA-KKAKOKEEDRNNTYITKLELEIA 216
DB 121 STKPSASTAKPNKTEPGEKVAEAKKVEEA-KKAKOKEEDRNNTYITKLELEIA 180
QY 217 ESDVVKAELELVKANKANPRDEOKIKQAEAVESKQAEATRLKIKTDREBEAEBAKR 276
DB 181 ESDVVKAELELVKANKANPRDEOKIKQAEAVESKQAEATRLKIKTDREBEAEBAKR 240
QY 277 RADAKEQKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPSPLKPEKKAFAEAK 336
DB 241 RADAKEQKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPSPLKPEKKAFAEAK 300
QY 337 KVEEKKKAEDQKEDDRNYPNTYKTELEIAESDVVKAELELVKAEKKEPRNEKV 396
DB 301 KVEEKKKAEDQKEDDRNYPNTYKTELEIAESDVVKAELELVKAEKKEPRNEKV 360
QY 397 KQAAVEESKKAETRLKIKTDREBEAEBAKRKAEBEDKVEKPAEQOPAPAKAEKP 456
DB 361 KQAAVEESKKAETRLKIKTDREBEAEBAKRKAEBEDKVEKPAEQOPAPAKAEKP 420
QY 457 AAPKPEPAPQPAEKFPADQAEEDVARRSEEEYNRLTQOQPKTEKPAQSTP 511
DB 421 AAPKPEPAPQPAEKFPADQAEEDVARRSEEEYNRLTQOQPKTEKPAQSTP 475

RESULT 5
US-09-056-019-38
; Sequence 38, Application US/09056019A
; Publication No. US20030096950A1
; GENERAL INFORMATION:
; APPLICANT: Tuomanen, Elaine I
; APPLICANT: Wizemann, Theresa
; APPLICANT: Masure, H. R.
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Koenig, Scott
; TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL
```

```
; TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED
; TITLE OF INVENTION: THEREFROM AND USES THEREOF
; FILE REFERENCE: 1340-1-017
; CURRENT APPLICATION NUMBER: US/09/056,019A
; CURRENT FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 38
; LENGTH: 460
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-056-019-38

Query Match      88.8%; Score 2252.5; DB 11; Length 460;
Best Local Similarity 99.6%; Pred. No. 8.6e-107;
Matches 457; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

QY 24 VVASLWGSVVAHATENEGATQVPTSSNRANESQAEQEQPKLDSERDKARKEVEYVK 83
DB 2 IVASLWGSVVAHATENEGATQVPTSSNRANESQAEQEQPKLDSERDKARKEVEYVK 61
QY 84 IVGESYAKSTKRRHTITVALVNELNINKIYKIVESTSESQLOIIMESRSKYDEAVS 143
DB 62 IVGESYAKSTKRRHTITVALVNELNINKIYKIVESTSESQLOIIMESRSKYDEAVS 121
QY 144 KFEKSSSSSSSDSTPEASDTAKPKPTPEGKVAEAKKVEEA-KKAKOKEEDRN 202
DB 122 KFEKSSSSSSSDSTPEASDTAKPKPTPEGKVAEAKKVEEA-KKAKOKEEDRN 181
QY 203 YPTITTYTLELEIAESDVVKAELELVKANKANPRDEOKIKQAEAVESKQAEATRLK 262
DB 182 YPTITTYTLELEIAESDVVKAELELVKANKANPRDEOKIKQAEAVESKQAEATRLK 241
QY 263 IKTDREBEAEBAKRADAKQKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPS 322
DB 242 IKTDREBEAEBAKRADAKQKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPS 301
QY 323 PSLKPEKKAFAEAKKVEEAKKKAEDQKEDDRNYPNTYITKLELEIAESDVVKAELEL 382
DB 302 PSLKPEKKAFAEAKKVEEAKKKAEDQKEDDRNYPNTYITKLELEIAESDVVKAELEL 361
QY 383 VKERAKPRNEKVKQAKAVESKKAETRLKIKTDREBEAEBAKRKAEBEDKVEKPA 442
DB 362 VKERAKPRNEKVKQAKAVESKKAETRLKIKTDREBEAEBAKRKAEBEDKVEKPA 421
QY 443 EQPAPAPAKAEKPAPAKPEPAPAEQPKAEKPADQAE 481
DB 422 EQPAPAPAKAEKPAPAKPEPAPAEQPKAEKPADQAE 460

RESULT 6
US-09-056-019-39
; Sequence 39, Application US/09056019A
; Publication No. US20030096950A1
; GENERAL INFORMATION:
; APPLICANT: Tuomanen, Elaine I
; APPLICANT: Wizemann, Theresa
; APPLICANT: Masure, H. R.
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Koenig, Scott
; TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL
; TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED
; FILE REFERENCE: 1340-1-017
; CURRENT APPLICATION NUMBER: US/09/056,019A
; CURRENT FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 39
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-056-019-39
```

Query Match 88.1%; Score 2235.5; DB 11; Length 459;
Best Local Similarity 98.7%; Pred. No. 6.2e-106;
Matches 453; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 24 VVASLWMSVYATNEGATVPPTSSNRANESQAEOGEPKLDSEBDKARKEVEEYK 83
DB 1 IVASLWMSVYATNEGATVPPTSSNRANESQAEOGEPKLDSEBDKARKEVEEYK 60

QY 84 IVGESYAKSTKKRHTITVALVNLNINKNEYLNKIVESTESQOILMMESRSKYDAVS 143
DB 61 IVGESYAKSTKKRHTITVALVNLNINKNEYLNKIVESTESQOILMMESRSKYDAVS 120

QY 144 KPEKSSSSSSSDSTKPEASDTAKPNKPTPEGEVVAEKKKVEEA-KKAKDQKEEDRN 202
DB 121 KPEKSSSSSSSDSTKPEASDTAKPNKPTPEGEVVAEKKKVEEKKADQKEEDRN 180

QY 203 YPTITTKLELEIASDVEVKAELELVKVKANEBRDEOKIKQAEAVESKQAEATRLK 262
DB 181 YPTITTKLELEIASDVEVKAELELVKVKANEBRDEOKIKQAEAVESKQAEATRLK 240

QY 263 IKTDREAEAEKRRADAKEQCKPKGRKGVPGELATPDKKENDAKSSDSSVGEETLPS 322
DB 241 IKTDREAEAEKRRADAKEQCKPKGRKGVPGELATPDKKENDAKSSDSSVGEETLPS 300

QY 323 PSLEKEKVAEAEKVEEAKKAEKEDRKNYPTNTYKLELEIASDVEVKAELEL 382
DB 301 PSLEKEKVAEAEKVEEAKKAEKEDRKNYPTNTYKLELEIASDVEVKAELEL 360

QY 383 VKEAEKERNNEKVAQAEVESKKAETRLKIKTDREKAEAEKRAEEDKYKERA 442
DB 361 VKEAEKERNNEKVAQAEVESKKAETRLKIKTDREKAEAEKRAEEDKYKERA 420

QY 443 EOPQAPAPKAEKPAAPKPPENPAEQPAEKPADQAE 481
DB 421 EOPQAPAPKAEKPAAPKPPENPAEQPAEKPADQAE 459

RESULT 7
US-10-254-995-9
; Sequence 9, Application US/10254995
; Publication No. US20030138447A1
; GENERAL INFORMATION:
; APPLICANT: Witzemann, Theresa M.
; APPLICANT: Koenig, Scott
; APPLICANT: Johnson, Leslie S
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
; FILE REFERENCE: 469201-396
; CURRENT APPLICATION NUMBER: US/10/254,995
; CURRENT FILING DATE: 2002-09-25
; PRIOR APPLICATION NUMBER: US/09/286,981
; PRIOR FILING DATE: 1999-04-06
; PRIOR APPLICATION NUMBER: US 60/085,743
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-254-995-9

Query Match 86.5%; Score 2194.5; DB 12; Length 446;
Best Local Similarity 99.8%; Pred. No. 7.2e-104;
Matches 445; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 37 TENEGATVPPTSSNRANESQAEOGEPKLDSEBDKARKEVEEYKIVGESYAKSTKR 96
DB 1 TENEGATVPPTSSNRANESQAEOGEPKLDSEBDKARKEVEEYKIVGESYAKSTKR 60

QY 97 HTITVALVNLNINKNEYLNKIVESTESQOILMMESRSKYDAVSKFEKSSSSSSSD 156
DB 61 HTITVALVNLNINKNEYLNKIVESTESQOILMMESRSKYDAVSKFEKSSSSSSSD 120

QY 157 SSTPEASDTAKPNKPTPEGEKVAEAKKKVEEA-KKAKDQKEEDRNPTITTKLELEI 215
DB 121 SSTPEASDTAKPNKPTPEGEKVAEAKKKVEEA-KKAKDQKEEDRNPTITTKLELEI 180

QY 216 AESDVEVKAELELVKVKANEBRDEOKIKQAEAVESKQAEATRLKIKTDREAEAEK 275
DB 181 AESDVEVKAELELVKVKANEBRDEOKIKQAEAVESKQAEATRLKIKTDREAEAEK 240

QY 276 RRADAKEQCKPKGRKGVPGELATPDKKENDAKSSDSSVGEETLPSPSLKEKKVAE 335
DB 241 RRADAKEQCKPKGRKGVPGELATPDKKENDAKSSDSSVGEETLPSPSLKEKKVAE 300

QY 336 KKVVEAKKKAADQKEEDRNPTNTYKLELEIASDVEVKAELELVKKEAEKERNNEK 395
DB 301 KKVVEAKKKAADQKEEDRNPTNTYKLELEIASDVEVKAELELVKKEAEKERNNEK 360

QY 396 VKQAEAVESKKAETRLKIKTDREKAEAEKRAEEDKYKERAPOPAAPAKAE 455
DB 361 VKQAEAVESKKAETRLKIKTDREKAEAEKRAEEDKYKERAPOPAAPAKAE 420

QY 456 PAPAPKPPENPAEQPAEKPADQAE 481
DB 421 PAPAPKPPENPAEQPAEKPADQAE 446

RESULT 8
US-10-254-995-6
; Sequence 6, Application US/10254995
; Publication No. US20030138447A1
; GENERAL INFORMATION:
; APPLICANT: Witzemann, Theresa M.
; APPLICANT: Koenig, Scott
; APPLICANT: Johnson, Leslie S
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
; FILE REFERENCE: 469201-396
; CURRENT APPLICATION NUMBER: US/10/254,995
; CURRENT FILING DATE: 2002-09-25
; PRIOR APPLICATION NUMBER: US/09/286,981
; PRIOR FILING DATE: 1999-04-06
; PRIOR APPLICATION NUMBER: US 60/085,743
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-254-995-6

Query Match 85.8%; Score 2177.5; DB 12; Length 446;
Best Local Similarity 98.9%; Pred. No. 5.2e-103;
Matches 441; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 37 TENEGATVPPTSSNRANESQAEOGEPKLDSEBDKARKEVEEYKIVGESYAKSTKR 96
DB 1 TENEGATVPPTSSNRANESQAEOGEPKLDSEBDKARKEVEEYKIVGESYAKSTKR 60

QY 97 HTITVALVNLNINKNEYLNKIVESTESQOILMMESRSKYDAVSKFEKSSSSSSSD 156
DB 61 HTITVALVNLNINKNEYLNKIVESTESQOILMMESRSKYDAVSKFEKSSSSSSSD 120

QY 157 SSTPEASDTAKPNKPTPEGEKVAEAKKKVEEA-KKAKDQKEEDRNPTITTKLELEI 215
DB 121 SSTPEASDTAKPNKPTPEGEKVAEAKKKVEEKKADQKEEDRNPTITTKLELEI 180

QY 216 AESDVEVKAELELVKVKANEBRDEOKIKQAEAVESKQAEATRLKIKTDREAEAEK 275
DB 181 AESDVEVKAELELVKVKANEBRDEOKIKQAEAVESKQAEATRLKIKTDREAEAEK 240

QY 276 RRADAKEQCKPKGRKGVPGELATPDKKENDAKSSDSSVGEETLPSPSLKEKKVAE 335
DB 241 RRADAKEQCKPKGRKGVPGELATPDKKENDAKSSDSSVGEETLPSPSLKEKKVAE 300

;; TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED
;; FILE OF INVENTION: THEREFOR AND USES THEREOF
;; FILE REFERENCE: 1340-1-017
;; CURRENT APPLICATION NUMBER: US/09/056,019A
;; CURRENT FILING DATE: 1998-04-07
;; NUMBER OF SEQ ID NOS: 39
;; SOFTWARE: Patent In Ver. 2.0
;; SEQ ID NO 1
;; LENGTH: 406
;; TYPE: PRT
;; ORGANISM: Streptococcus pneumoniae
US-09-056-019-1

Query Match 78.0%; Score 1979.5; DB 11; Length 406;
Best Local Similarity 99.8%; Pred. No. 4.9e-93;
Matches 405; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 38 ENEGATQVPTSSNRANESQAEQGPCKLDSERDARKVEEYVKKIVGESYAKSTKRKH 97
DB 1 ENEGATQVPTSSNRANESQAEQGPCKLDSERDARKVEEYVKKIVGESYAKSTKRKH 60
QY 98 TTTVALVNLNNIKNEYINKIVESTSESQLOLIMESRSKYDEAVSKTEKSSSSSSSDS 157
DB 61 TTTVALVNLNNIKNEYINKIVESTSESQLOLIMESRSKYDEAVSKTEKSSSSSSSDS 120
QY 158 STKPASTPAKPNKTEPEGEKVAEAKKVEEA-KKAKOKEDRRNYPITTYKTLELEIA 216
DB 121 STKPASTPAKPNKTEPEGEKVAEAKKVEEA-KKAKOKEDRRNYPITTYKTLELEIA 180
QY 217 ESDVEVKKAELVVKANEPREDEQIKQAEAEVSKQAEATRLKKITDREAEAEAKR 276
DB 181 ESDVEVKKAELVVKANEPREDEQIKQAEAEVSKQAEATRLKKITDREAEAEAKR 240
QY 277 RADAKEQCKPGRAGVPGELATPDKKENDAKSSDSSVGEETLPSLKPCKVAAEAK 336
DB 241 RADAKEQCKPGRAGVPGELATPDKKENDAKSSDSSVGEETLPSLKPCKVAAEAK 300
QY 337 KVEEAKKKAEDQKEDRRNYPITTYKTLELEIAESDVVKKAELVKEAEKPEPNEBKV 396
DB 301 KVEEAKKKAEDQKEDRRNYPITTYKTLELEIAESDVVKKAELVKEAEKPEPNEBKV 360
QY 397 KQAKAEVSKKAEATRLKIKTDRAKAEAEAKKAAEDDKVKEKA 442
DB 361 KQAKAEVSKKAEATRLKIKTDRAKAEAEAKKAAEDDKVKEKA 406

RESULT 12

US-09-298-523B-9
; Sequence 9, Application US/09298523B
; Publication No. US20030059438A1
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/298,523B
; CURRENT FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 9
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-298-523B-9

Query Match 69.4%; Score 1761.5; DB 11; Length 487;
Best Local Similarity 71.5%; Pred. No. 6.3e-82;
Matches 376; Conservative 44; Mismatches 51; Indels 55; Gaps 8;

QY 1 MFASKERKHYHSIRKFSVG-ASVVVASLWGSVVHATENEGATQVPTSSNRANESQAEQ 59
DB 1 MFASKERKHYHSIRKFSIGVAVASLWGSVVHATEKEVTTQVATSPFNANKSQ--- 57

QY 60 GEOPCKLDSERDARKVEEYVKKIVGESYAKSTKSHITVALVNLNNIKNEYINKIV 119
DB 58 -----TEHMKAKAQVDEYITTKL-----QDRRHGTQVGLITKGVIKTEYLRHS 104
QY 120 ESTSESQLOLIMESRSKYDEAVSKFEKSSSSSSSDSSTKPEASDTAKPNKPTPEGEKV 179
DB 105 VSKKESAE-LPSEIKAKLDAFQPKD-----LPTEPGKVV 142
QY 180 AEAKKVEEA-KKAKOKEDRRNYPITTYKTLELEIAESDVVKKAELVVKANEP 238
DB 143 AEAKKVEEA-KKAKOKEDRRNYPITTYKTLELEIAESDVVKKAELVVKANEP 202
QY 239 DEQIKQAEAEVSKQAEATRLKKITDREAE-----EAKRRADAKEQ 284
DB 203 NEQVNOAKAVESKQAEATRLKKITDREAEATRLLENITKDREKAEAKRKADQD 262
QY 285 KPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPSLKPCKVAAEAKKVEEAKK 344
DB 263 ESKRRVKGVPQATLDPKKENDAKSSDSSVGEETLPSLKSQKVAEAKKVAEAKK 322
QY 345 AEDQKEDRRNYPITTYKTLELEIAESDVVKKAELVKEAEKPEPNEBKVKAKEVE 404
DB 323 AKDQKEDRRNYPITTYKTLELEIAESDVVKKAELVKEAEKPEPNEBKVKAKEVE 382
QY 405 SKKAEATRLKIKTDRAKAEAEAKKAAEDDKVKEKPAEQPPAPAPAEKPAKPEP 464
DB 383 SKKAEATRLKIKTDRAKAEAEAKKAAEDDKVKEKPAEQPPAPAPAEKPAKPEP 441
QY 465 PAEQPKAEKPADQQAEDDYARSEEYNRLLTQQQPPTEKPAQST 510
DB 442 PAEQPKAEKPADQQAEDDYARSEEYNRLLTQQQPPTEKPAQST 487

RESULT 13

US-09-298-523B-66
; Sequence 66, Application US/09298523B
; Publication No. US20030059438A1
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/298,523B
; CURRENT FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 66
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-298-523B-66

Query Match 68.8%; Score 1745.5; DB 11; Length 487;
Best Local Similarity 71.3%; Pred. No. 4.1e-81;
Matches 376; Conservative 45; Mismatches 49; Indels 57; Gaps 10;

QY 1 MFASKERKHYHSIRKFSVG-ASVVVASLWGSVVHATENEGATQVPTSSNRANESQAEQ 59
DB 1 MFASKERKHYHSIRKFSIGVAVASLWGSVVHATEKEVTTQVATSPFNANKSQ--- 57
QY 60 GEOPCKLDSERDARKVEEYVKKIVGESYAKSTKSHITVALVNLNNIKNEYINKIV 119
DB 58 -----TEHMKAKAQVDEYITTKL-----QDRRHGTQVGLITKGVIKTEYLRHS 104
QY 120 ESTSESQLOLIMESRSKYDEAVSKFEKSSSSSSSDSSTKPEASDTAKPNKPTPEGEKV 179
DB 105 VSKKESAE-LPSEIKAKLDAFQPKD-----LPTEPGKVV 142
QY 180 AEAKKVEEA-KKAKOKEDRRNYPITTYKTLELEIAESDVVKKAELVVKANEP 238
DB 143 AEAKKVEEA-KKAKOKEDRRNYPITTYKTLELEIAESDVVKKAELVVKANEP 202
QY 239 DEQIKQAEAEVSKQAEATRLKKITDREAE-----EAKRRADAKE-Q 283

Db	203	NEQVKNQAKAVESQOAEATRLKIKITDREQOAEATRLINIKITDRREKAEEARKKAAEAIVK	262
Qy	284	GKPKRARAGVGBELATPDDKKNDAKSSDSVSGEETLTSPSLIKPEKKYVAEAKKYVEAAK	343
Db	263	DLKRLRTGAAPGEPAITPDKKNDAKSSDSVSGEETLTSPSLIKSGKVAEAKKYAAEK	322
Qy	344	KAEDQKEEDRNRYPTNTYKLTLELEIAESDVEYKKELEIVKEAKEPNEEKVKQAKAEV	403
Db	323	KAKQKEDERNRYPTNTYKLTLELE-AESDVKYKEALEIVKEAASESNEEKVKQAKAEV	381
Qy	404	ESKKAELTRLEKITDRKKAEEARKKAAEEDYKVAEAPQOPAPAPAEKPPAPAPAE	463
Db	382	ESKKAELTRLEKITDRKKA-EEARKKAAEEDYKVAEAPQOPAPAPQEPKAPAPAE	440
Qy	464	NPAPQPAKEKPADQOAEEDYARSRSEENRRLTQOQPPTEKPAQOST	510
Db	441	NPAPQPAKEKPADQOAEEDYARSRSEENRRLTQOQPPTEKPAQOST	487

```

RESULT 14
US-09-298-523B-56
: Sequence 56, Application US/092986523B
: Publication No. US20030059438A1
: GENERAL INFORMATION:
: APPLICANT: BRILES et al.
: TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
: FILE OF INVENTION: AND STRAINS THEREOF AND USES THEREFOR
: FILE REFERENCE: 454312-3140
: CURRENT APPLICATION NUMBER: US/09/298,523B
: CURRENT FILING DATE: 1999-04-23
: NUMBER OF SEQ ID NOS: 78
: SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 56
LENGTH: 581
TYPE: PRT
: ORGANISM: Streptococcus pneumoniae
US-09-298-523B-56

```

	Query Match	67.7%;	Score 1718;	DB 11;	Length 581;	
	Best Local Similarity	70.3%;	Pred. No. 1.2e-79;			
	Matches	369;	Conservative	49;	Mismatches	44;
					Indels	58;
					Gaps	10;
QY	1	MFASKSERKHYVSI	RKFSVG--ASVVAVSLVMSGVVHATENEGATQVPTSSNRANESQAQ	59		
DB	1	MFASKSERKHYVSI	RKFSIGVASVAVSLVMSGVVHATENEGSTQATISNMAK-----	54		
QY	60	GEQFKLIDSEDDKARKE--	VEEYVAKKIYESASTKKRHITVALVNLNINKOEYLNK-	117		
DB	55	-----TEHRKAKAQV	VDVEYIEMLRE--TQDRRHGTGNVALNITKLSIAIKTYLREL	104		
QY	118	-IVESTSPSOQLIM	MSRSKYVDEVSKSEFEKDESSSSSSSDSSTPEASDTAKPKKPTMBG	176		
DB	105	NVLBEKSDLEP---	SEIKAKLIDAAFEKFKQOT-----LKPQ	138		
QY	177	EKVAEAKKVV	EEA--KKAKDQKEEDRRNPPTIYKTLLELIAESDVEVKAELVLYKVN	235		
DB	139	EKVAEAKKVV	EEAKKAKADQKEEDRRNPPTIYKTLLELIAEPVYKVAELVLYKEBAK	198		
QY	236	EPREDEKIKQAE	AEVSKQAEATLTUKKITDREAEAEFAKRADAK-----EQKP	286		
DB	199	ESRNEGTTKQAE	KEVESKQAEATLTENKITDRKAAEEFAKRAKADAKLEANVATSDQKP	258		
QY	287	KGRAKRGV	GPGLATPPDKKENAKSSDSVGEETLPSPSLKPEKKVVAEAKKVEAEAKKAA	346		
DB	259	KGRAKRGV	GPGLATPPDKKENAKSSDSVGEETLPSSSLKGGKVAEAEKKVVEAEAKKAA	318		
QY	347	DQKEEDRRNP	PTNTYKTLLELIAESDVEVKAELVLYKEAEPRNEBKVQKAAAEVSK	406-		
DB	319	DQKEEDRRNP	PTNTYKTLLELIAESDVKVKAELVLYKEAEPRDEBKIKQAAKAEVSK	378		
QY	407	KAEATREL	EKIKTDPKAAEEFAKRAEEDKYKEKPAQPPQAPAPKAEKPPAKPENPA	466		

Db 379 KAEATRLLENITKTRKKAAEEAKRKRAAAEDYKAEKABEQPPAPATQPEK--PAKPEKPA 436

QY 467 EQPKAEKPADQQAEDYARSRSEETNRLTQQQPPKTEKAQGSTP 511

Db 437 EQPKAEKTDQQAEDYARSRSEETNRLTQQQPPKTEKAQGFPT 481

```

RESULT 15
US-09-298-523B-10
; Sequence 10, Application US/09298523B
; Publication No. US20030059438A1
; GENERAL INFORMATION:
; APPLICANT: BRILES et al
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; FILE OF INVENTION: AND STRAINS THEROPF AND USES THEREFOR
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/298,523B
; CURRENT FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-298-523B-10

```

Query Match	Best Local Similarity	70.6%	Score 1715;	DB 11;	Length 483;
Matches	369;	Conservative	48;	Mismatches	50;
				Indels	58;
				Gaps	10;
Qy	1	MFASKSRKXVHYSIRKFSVG-ASVVAASVLVMSGVVHAHATENEGATQVPTSSRNANESQAO	59		
Db	1	MFASKSRKXVHYSIRKFSIGVASVAASVLVMSGVVHAHATENEGSTQATSSNMAK-----	54		
Qy	60	GEQPKLDSERDKARKE-VEEYKXKLVGESYAKSTKRGHITTVLVNVEINNIRKYEVLNK-	117		
Db	55	-----THRRKAQKVDEYIEKMLRE--IQDRKRHHQVNAVLNKLGAIKKTYLREL	104		
Qy	118	-IYVTSSESQQLIMMSRSKVDDEAVSGFPEKSSSSSSSSSSSTKPEASDTKPKKPTPEPG	176		
Db	105	NVLBEKSKDEL--SEIKAKLDIAFFKFKDT-----LKPG	138		
Qy	177	EKVAEAKKKVVEA-KKAKDQKEEDRRNYPITYKTLELEIAESDVEYKAELELVKANK	235		
Db	139	EKVAEAKKKVVEAKKKAKDQKEEDRRNYPITYKTLELEIAEPVYKXKAELELVKAEAK	198		
Qy	236	EPPEQKIKQAEAVESKQAEATRLKKIKTIDREBAEEAKRRADAK-----EGCKP	286		
Db	199	ESRENGTIIKQAKEVESKKAEATRLENIKTDRKKAEEBAERKADAKLKEANVATSDQKP	258		
Qy	287	KGRAKRGVPBELATPPDKENDAKSSDSVGEETLPSSSLPKEKKVAEAKKVEAKKAAE	346		
Db	259	KGRAKRGVPBELATPPDKENDAKSSDSVGEETLPSSSLSGKKVVAEAKKVEAEKKA	318		
Qy	347	DOKEEDRRNPTNTYKTLLEIAESDVEYKAELELVKEAKEPRNEKXKQAKAEVSK	406		
Db	319	DOKEEDRRNPTNTYKTLLEIAESDVKAELELVKEAKEPRDEKTIKQAKAEVSK	378		
Qy	407	KAETATRLKIKTDRKKAEEBAKRAAEEDVYKEKPAEQOPAPAPAKAEKAPAPKPPENPA	466		
Db	379	KAETATRLKIKTDRKKAEEBAKRAAEEDVYKEKPAEQOPAPATQEEK--PAKPEKPA	436		
Qy	467	EQRKAKEPADQQAEBDYARSEEYNNRLTQOQPPKTEKPAQPSIF	511		
Db	437	EQRKAKEPTDQQAEBDYARSEEYNNRLTQOQPPKTEKPAQPSIF	481		

Search completed: November 21, 2003, 13:42:42
Job time : 40.1939 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:33:52 ; Search time 13.8061 Seconds

(without alignments)
2380.162 Million cell updates/sec

Title: US-09-298-523C-13_COPY_263_442

Perfect score: 889

Sequence: 1 IKTDREAEAEKRRADAKE.....AEEAKKKAEDKVEKKA 180

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubppaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubppaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubppaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubppaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubppaa/US09C_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubppaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubppaa/US10C_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubppaa/US10C_PUBCOMB.pep.*
18: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	889	100.0	406 11 US-09-056-019-1	Sequence 1, Appl1
2	889	100.0	428 11 US-09-056-019-24	Sequence 24, Appl1
3	889	100.0	446 12 US-10-254-995-9	Sequence 9, Appl1
4	889	100.0	453 9 US-09-765-272-38	Sequence 38, Appl1
5	889	100.0	460 11 US-09-056-019-38	Sequence 13, Appl1
6	889	100.0	511 11 US-09-298-523B-13	Sequence 13, Appl1
7	889	100.0	513 11 US-09-298-523B-12	Sequence 12, Appl1
8	889	100.0	655 11 US-09-056-019-2	Sequence 2, Appl1
9	889	100.0	693 11 US-09-769-787-185	Sequence 185, App
10	884	99.4	446 12 US-10-254-995-6	Sequence 6, Appl1
11	884	99.4	459 11 US-09-056-019-39	Sequence 39, Appl1
12	861	96.9	451 11 US-09-298-523B-67	Sequence 67, Appl1
13	803.5	90.4	376 11 US-09-056-019-7	Sequence 7, Appl1
14	803.5	90.4	413 11 US-09-056-019-35	Sequence 35, Appl1
15	803.5	90.4	414 12 US-10-254-995-16	Sequence 16, Appl1

16	803.5	90.4	581 11 US-09-298-523B-56	Sequence 56, Appl1
17	803.5	90.4	631 12 US-09-829-382-25	Sequence 25, Appl1
18	803.5	90.4	663 11 US-09-298-523B-58	Sequence 58, Appl1
19	803.5	90.4	663 11 US-09-056-019-8	Sequence 8, Appl1
20	803.5	90.4	663 12 US-09-969-748C-14	Sequence 14, Appl1
21	803.5	90.4	663 12 US-09-949-039-104	Sequence 104, App
22	800.5	90.0	483 11 US-09-298-523B-10	Sequence 10, Appl1
23	795	89.4	439 11 US-09-056-019-30	Sequence 30, Appl1
24	793	89.2	419 11 US-09-056-019-31	Sequence 31, Appl1
25	793	89.2	444 11 US-09-298-523B-5	Sequence 5, Appl1
26	792	89.1	406 12 US-10-254-995-18	Sequence 18, Appl1
27	792	89.1	425 12 US-10-254-995-11	Sequence 11, Appl1
28	791.5	89.0	414 12 US-10-254-995-17	Sequence 17, Appl1
29	789	88.8	481 11 US-09-298-523B-6	Sequence 6, Appl1
30	784	88.2	428 12 US-10-254-995-7	Sequence 7, Appl1
31	784	88.2	439 11 US-09-056-019-37	Sequence 37, Appl1
32	784	88.2	496 11 US-09-298-523B-4	Sequence 4, Appl1
33	777.5	87.5	437 11 US-09-298-523B-64	Sequence 64, Appl1
34	775.5	87.2	412 12 US-10-254-995-17	Sequence 17, Appl1
35	775.5	87.2	425 11 US-09-056-019-36	Sequence 36, Appl1
36	771.5	86.8	439 11 US-09-056-019-28	Sequence 28, Appl1
37	770.5	86.7	424 12 US-10-254-995-14	Sequence 14, Appl1
38	770.5	86.7	426 12 US-10-254-995-12	Sequence 12, Appl1
39	770.5	86.7	437 11 US-09-056-019-29	Sequence 29, Appl1
40	769.5	86.6	425 12 US-10-254-995-13	Sequence 13, Appl1
41	769.5	86.6	437 11 US-09-056-019-32	Sequence 32, Appl1
42	768	86.4	419 12 US-10-254-995-15	Sequence 15, Appl1
43	768	86.4	433 11 US-09-056-019-33	Sequence 33, Appl1
44	751.5	84.5	427 11 US-09-056-019-34	Sequence 34, Appl1
45	743	83.6	487 11 US-09-298-523B-9	Sequence 9, Appl1

ALIGNMENTS

RESULT 1
US-09-056-019-1
Sequence 1, Application US/09056019A
Publication No. US20030096950A1
GENERAL INFORMATION:
APPLICANT: Tuomann, Elaine I
APPLICANT: Witzmann, Theresa
APPLICANT: Measure, H. R.
APPLICANT: Johnson, Leslie S.
TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL
TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED
FILE REFERENCE: 1340-1-017
CURRENT APPLICATION NUMBER: US/09/056,019A
CURRENT FILING DATE: 1998-04-07
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 406
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-056-019-1

Query Match 100.0%; Score 889; DB 11; Length 406;
Best Local Similarity 100.0%; Pred. No. 5.1e-53;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IKTDREAEAEKRRADAKEQKPKRAGVGEIATPDKKENDAKSSDSVGEETLPS 60
Db IKTDREAEAEKRRADAKQKPKRAGVGEIATPDKKENDAKSSDSVGEETLPS 266
QY 61 PSLKPEKVAEAKKYEAEKKAEDQKEBRNRYPTNTYTTLELAESVEVKKLEL 120
Db 287 PSLKPEKVAEAKKYEAEKKAEDQKEBRNRYPTNTYTTLELAESVEVKKLEL 346
QY 121 VKKEAEPRREEVVKAKAEVSKKAEATLEKIKTDKKAEAEAKKAAEPRVKEKPA 180

Db 347 VKEAKEPRNEEKVQAKAEVESKKAATRLLEKIKTRDKKAAEEBAKRAAEEDVKKEKPA 406

RESULT 2

US-09-056-019-24

Sequence 24, Application US/09056019A

Publication No. US20030096950A1

GENERAL INFORMATION:

APPLICANT: Tuomanen, Elaine I

APPLICANT: Wizemann, Theresa

APPLICANT: Masure, H. R.

APPLICANT: Johnson, Leslie S.

APPLICANT: Koenig, Scott

TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL

TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED

FILE REFERENCE: 1340-1-017

CURRENT APPLICATION NUMBER: US/09/056,019A

CURRENT FILING DATE: 1998-04-07

NUMBER OF SEQ ID NOS: 39

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 24

LENGTH: 428

TYPE: PRT

ORGANISM: Streptococcus pneumoniae

US-09-056-019-24

Query Match 100.0%; Score 889; DB 11; Length 428;

Best Local Similarity 100.0%; Pred. No. 5,4e-53;

Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IKTDREAEBAEKRRADAKGQKPKGRKRGVPGELATPPDKKENDAKSSDSSVGEETLPS 60

Db 227 IKTDREAEBAEKRRADAKGQKPKGRKRGVPGELATPPDKKENDAKSSDSSVGEETLPS 286

Qy 61 PSLKPEKKVAEAEKKVEAKKKAEDQKEEDRRNYPNTTYKTLELEIAESDVEVKAELEL 120

Db 287 PSLKPEKKVAEAEKKVEAKKKAEDQKEEDRRNYPNTTYKTLELEIAESDVEVKAELEL 346

Qy 121 VKEAKEPRNEEKVQAKAEVESKKAATRLLEKIKTRDKKAAEEBAKRAAEEDVKKEKPA 180

Db 347 VKEAKEPRNEEKVQAKAEVESKKAATRLLEKIKTRDKKAAEEBAKRAAEEDVKKEKPA 406

RESULT 3

US-10-254-995-9

Sequence 9, Application US/10254995

Publication No. US20030138447A1

GENERAL INFORMATION:

APPLICANT: Wizemann, Theresa M.

APPLICANT: Koenig, Scott

APPLICANT: Johnson, Leslie S

TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines

FILE REFERENCE: 469201-396

CURRENT APPLICATION NUMBER: US/10/254,995

CURRENT FILING DATE: 2002-09-25

PRIOR APPLICATION NUMBER: US/09/286,981

PRIOR FILING DATE: 1999-04-06

PRIOR APPLICATION NUMBER: US 60/085,743

PRIOR FILING DATE: 1998-05-15

NUMBER OF SEQ ID NOS: 38

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 9

LENGTH: 446

TYPE: PRT

ORGANISM: Streptococcus pneumoniae

US-10-254-995-9

Query Match 100.0%; Score 889; DB 12; Length 446;

Best Local Similarity 100.0%; Pred. No. 5,7e-53;

Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IKTDREAEBAEKRRADAKGQKPKGRKRGVPGELATPPDKKENDAKSSDSSVGEETLPS 60

Db 228 IKTDREAEBAEKRRADAKGQKPKGRKRGVPGELATPPDKKENDAKSSDSSVGEETLPS 287

Qy 61 PSLKPEKKVAEAEKKVEAKKKAEDQKEEDRRNYPNTTYKTLELEIAESDVEVKAELEL 120

Db 288 PSLKPEKKVAEAEKKVEAKKKAEDQKEEDRRNYPNTTYKTLELEIAESDVEVKAELEL 347

Qy 121 VKEAKEPRNEEKVQAKAEVESKKAATRLLEKIKTRDKKAAEEBAKRAAEEDVKKEKPA 180

Db 348 VKEAKEPRNEEKVQAKAEVESKKAATRLLEKIKTRDKKAAEEBAKRAAEEDVKKEKPA 407

RESULT 4

US-09-765-272-38

Sequence 38, Application US/09765272

Patent No. US20020061545A1

GENERAL INFORMATION:

APPLICANT: Choi et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines

NUMBER OF SEQUENCES: 452

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 Inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/765,272

FILING DATE: 22-Jan-2001

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/961,083

FILING DATE: <unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 38:

SEQUENCE CHARACTERISTICS:

LENGTH: 453 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 38:

US-09-765-272-38

Query Match 100.0%; Score 889; DB 9; Length 453;

Best Local Similarity 100.0%; Pred. No. 5,7e-53;

Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IKTDREAEBAEKRRADAKGQKPKGRKRGVPGELATPPDKKENDAKSSDSSVGEETLPS 60

Db 25 IKTDREAEBAEKRRADAKGQKPKGRKRGVPGELATPPDKKENDAKSSDSSVGEETLPS 84

Qy 61 PSLKPEKKVAEAEKKVEAKKKAEDQKEEDRRNYPNTTYKTLELEIAESDVEVKAELEL 120

Db 85 PSLKPEKKVAEAEKKVEAKKKAEDQKEEDRRNYPNTTYKTLELEIAESDVEVKAELEL 144

Qy 121 VKEAKEPRNEEKVQAKAEVESKKAATRLLEKIKTRDKKAAEEBAKRAAEEDVKKEKPA 180

Db 145 VKEAKEPRNEEKVQAKAEVESKKAATRLLEKIKTRDKKAAEEBAKRAAEEDVKKEKPA 204

```
RESULT 5
US-09-056-019-38
; Sequence 38, Application US/09056019A
; Publication No. US20030096950A1
; GENERAL INFORMATION:
; APPLICANT: Tuomanen, Elaine I
; APPLICANT: Wazemann, Theresa
; APPLICANT: Masure, H. R.
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Koenig, Scott
; TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL
; TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED
; FILE REFERENCE: 1340-1-017
; CURRENT APPLICATION NUMBER: US/09/056,019A
; CURRENT FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 38
; LENGTH: 460
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-056-019-38

Query Match
Best Local Similarity 100.0%; Score 889; DB 11; Length 460;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPS 60
Db 242 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPS 301
Qy 61 PSLKPEKKVAEAEKKVEAKKKAEDQKEEDRRNPNTYTKTLELEIASDVEVKKAELEL 120
Db 302 PSLKPEKKVAEAEKKVEAKKKAEDQKEEDRRNPNTYTKTLELEIASDVEVKKAELEL 361
Qy 121 VKEAKEPRNEEKYQKAAEVESKKAETRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 180
Db 362 VKEAKEPRNEEKYQKAAEVESKKAETRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 421

RESULT 6
US-09-298-523B-13
; Sequence 13, Application US/09298523B
; Publication No. US20030059438A1
; GENERAL INFORMATION:
; APPLICANT: Briles et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; TITLE OF INVENTION: AND STRAINS THEREOF AND USES THEREFOR
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/298,523B
; CURRENT FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 13
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-298-523B-13

Query Match
Best Local Similarity 100.0%; Score 889; DB 11; Length 511;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPS 60
Db 263 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPS 322
Qy 61 PSLKPEKKVAEAEKKVEAKKKAEDQKEEDRRNPNTYTKTLELEIASDVEVKKAELEL 120
Db 323 PSLKPEKKVAEAEKKVEAKKKAEDQKEEDRRNPNTYTKTLELEIASDVEVKKAELEL 382
Qy 121 VKEAKEPRNEEKYQKAAEVESKKAETRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 180
Db 121 VKEAKEPRNEEKYQKAAEVESKKAETRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 180
```

```
Db 383 VKEAKEPRNEEKYQKAAEVESKKAETRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 442

RESULT 7
US-09-298-523B-12
; Sequence 12, Application US/09298523B
; Publication No. US20030059438A1
; GENERAL INFORMATION:
; APPLICANT: Briles et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; TITLE OF INVENTION: AND STRAINS THEREOF AND USES THEREFOR
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/298,523B
; CURRENT FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 12
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-298-523B-12

Query Match
Best Local Similarity 100.0%; Score 889; DB 11; Length 513;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPS 60
Db 263 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPS 322
Qy 61 PSLKPEKKVAEAEKKVEAKKKAEDQKEEDRRNPNTYTKTLELEIASDVEVKKAELEL 120
Db 323 PSLKPEKKVAEAEKKVEAKKKAEDQKEEDRRNPNTYTKTLELEIASDVEVKKAELEL 382
Qy 121 VKEAKEPRNEEKYQKAAEVESKKAETRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 180
Db 383 VKEAKEPRNEEKYQKAAEVESKKAETRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 442

RESULT 8
US-09-056-019-2
; Sequence 2, Application US/09056019A
; Publication No. US20030096950A1
; GENERAL INFORMATION:
; APPLICANT: Tuomanen, Elaine I
; APPLICANT: Wazemann, Theresa
; APPLICANT: Masure, H. R.
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Koenig, Scott
; TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL
; TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED
; FILE REFERENCE: 1340-1-017
; CURRENT APPLICATION NUMBER: US/09/056,019A
; CURRENT FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 655
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-056-019-2

Query Match
Best Local Similarity 100.0%; Score 889; DB 11; Length 655;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPS 60
Db 227 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPS 286
Qy 61 PSLKPEKKVAEAEKKVEAKKKAEDQKEEDRRNPNTYTKTLELEIASDVEVKKAELEL 120
```

; SEQ ID NO 67
; LENGTH: 451

TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-298-523B-67

Query Match 96.9%; Score 861; DB 11; Length 451;
Best Local Similarity 95.7%; Pred. No. 4,6e-51;
Matches 179; Conservative 0; Mismatches 0; Indels 8; Gaps 2;

QY 1 IKTDREAEAEAKRRADAK-----EOKPKGRAGVPGELATPDKKENDAKSSDSVGEETLPS 60
DB 195 IKTDREAEAEAKRRADAK-----EOKPKGRAGVPGELATPDKKENDAKSSDSVGEETLPS 254
QY 61 PSLAKEKVAEAEKVEAEKKAEADQKEDRRNYPNTYKTLLEIAESDVVKKAEL 120
DB 255 PSLAKEKVAEAEKVEAEKKAEADQKEDRRNYPNTYKTLLEIAESDVVKKAEL 313
QY 121 VKEAEKPRNEBKVKQA-----KAVESKKAETRLKIKTDRKKAEEAEKRAAEED 173
DB 314 VKEAEKPRNEBKVKQA-----KAVESKKAETRLKIKTDRKKAEEAEKRAAEED 373
QY 174 VKYKEKA 180
DB 374 VKYKEKA 380

RESULT 13

US-09-056-019-7
Sequence 7, Application US/09056019A
Publication No. US20030096950A1

GENERAL INFORMATION:
APPLICANT: Wizemann, Elaine I
APPLICANT: Wizemann, Theresa
APPLICANT: Masure, H. R.
APPLICANT: Johnson, Leslie S.
TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL
TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED
FILE REFERENCE: 1340-1-017
CURRENT APPLICATION NUMBER: US/09/056,019A
CURRENT FILING DATE: 1998-04-07
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO: 7
LENGTH: 376
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-056-019-7

Query Match 90.4%; Score 803.5; DB 11; Length 376;
Best Local Similarity 86.8%; Pred. No. 3e-47;
Matches 164; Conservative 12; Mismatches 4; Indels 9; Gaps 1;

QY 1 IKTDREAEAEAKRRADAK-----EOKPKGRAGVPGELATPDKKENDAKSSDS 51
DB 188 IKTDREAEAEAKRRADAK-----EOKPKGRAGVPGELATPDKKENDAKSSDS 247
QY 52 SVGEETLPSLKEKVAEAEKVEAEKKAEADQKEDRRNYPNTYKTLLEIAESDV 111
DB 248 SVGEETLPSLKEKVAEAEKVEAEKKAEADQKEDRRNYPNTYKTLLEIAESDV 307
QY 112 EVKAELELVKEAEKPRNEBKVKQA-----KAVESKKAETRLKIKTDRKKAEEAEKRAAE 171
DB 308 EVKAELELVKEAEKPRNEBKVKQA-----KAVESKKAETRLKIKTDRKKAEEAEKRAAE 367
QY 172 EDKYKEKA 180
DB 368 EDKYKEKA 376

RESULT 14
US-09-056-019-35
Sequence 35, Application US/09056019A

Publication No. US20030096950A1
GENERAL INFORMATION:
APPLICANT: Tuomanen, Elaine I
APPLICANT: Wizemann, Theresa
APPLICANT: Masure, H. R.
APPLICANT: Johnson, Leslie S.
APPLICANT: Koenig, Scott
TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL
TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED
FILE REFERENCE: 1340-1-017
CURRENT APPLICATION NUMBER: US/09/056,019A
CURRENT FILING DATE: 1998-04-07
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO: 35
LENGTH: 413
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-056-019-35

Query Match 90.4%; Score 803.5; DB 11; Length 413;
Best Local Similarity 86.8%; Pred. No. 3.3e-47;
Matches 164; Conservative 12; Mismatches 4; Indels 9; Gaps 1;

QY 1 IKTDREAEAEAKRRADAK-----EOKPKGRAGVPGELATPDKKENDAKSSDS 51
DB 188 IKTDREAEAEAKRRADAK-----EOKPKGRAGVPGELATPDKKENDAKSSDS 247
QY 52 SVGEETLPSLKEKVAEAEKVEAEKKAEADQKEDRRNYPNTYKTLLEIAESDV 111
DB 248 SVGEETLPSLKEKVAEAEKVEAEKKAEADQKEDRRNYPNTYKTLLEIAESDV 307
QY 112 EVKAELELVKEAEKPRNEBKVKQA-----KAVESKKAETRLKIKTDRKKAEEAEKRAAE 171
DB 308 EVKAELELVKEAEKPRNEBKVKQA-----KAVESKKAETRLKIKTDRKKAEEAEKRAAE 367
QY 172 EDKYKEKA 180
DB 368 EDKYKEKA 376

RESULT 15

US-10-254-995-16
Sequence 16, Application US/10254995
Publication No. US20030138447A1

GENERAL INFORMATION:
APPLICANT: Wizemann, Theresa M.
APPLICANT: Koenig, Scott
APPLICANT: Johnson, Leslie S.
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
FILE REFERENCE: 469201-396
CURRENT APPLICATION NUMBER: US/10/254,995
CURRENT FILING DATE: 2002-09-25
PRIOR APPLICATION NUMBER: US/09/266,981
PRIOR FILING DATE: 1999-04-06
PRIOR APPLICATION NUMBER: US 60/085,743
NUMBER OF SEQ ID NOS: 38
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO: 16
LENGTH: 414
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-10-254-995-16

Query Match 90.4%; Score 803.5; DB 12; Length 414;
Best Local Similarity 86.8%; Pred. No. 3.3e-47;
Matches 164; Conservative 12; Mismatches 4; Indels 9; Gaps 1;

QY 1 IKTDREAEAEAKRRADAK-----EOKPKGRAGVPGELATPDKKENDAKSSDS 51
DB 189 IKTDREAEAEAKRRADAK-----EOKPKGRAGVPGELATPDKKENDAKSSDS 248

QY	52	SVGEETLSPSPUKPEKVAEAEKKVBEAKKKAEDQKEDRRNYPNTYKTLLEIAESDV	111
Db	249	SVGEETLSPSSSLKSGKVAEAEKKVBEAKKKAEDQKEDRRNYPNTYKTLLEIAESDV	308
QY	112	EVKKALELVKEEAKPRNEEKVOKAAEVSKKAATRIEKITDRKKAEEEAARKAAE	171
Db	309	KVKEAELELVKEEAKPRDEEKIKOKAKVESKKAATRIENIKTDKKAEEEAARKAAE	368
QY	172	EDKVKKEKPA	180
Db	369	EDKVKKEKPA	377

Search completed: November 21, 2003, 13:42:43
Job time : 14.8061 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:23:06 ; Search time 14.7902 Seconds
(without alignments)
1624.773 Million cell updates/sec

Title: US-09-298-523C-13
Perfect score: 2537

Sequence: 1 MFASKSERKHYISIRKFSVG.....NRLTQQQPPKTEKAPQSTP 511

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	363	14.3	1020	NFH_HUMAN	P12036 homo sapien
2	351.5	13.9	1164	BAG_STRAG	P27951 streptococc
3	328	12.9	1391	MSTI_DROHY	Q08696 drosophila
4	328	12.9	6632	UN89_CAEEL	Q01761 caenorhabdi
5	316	12.5	831	NFH_RAT	P16884 rattus norv
6	315.5	12.4	705	TRDN_RABIT	Q28820 oryctolagus
7	311	12.3	539	M24_STRPY	P12379 streptococc
8	310	12.2	1087	NFH_MOUSE	P12946 mus musculu
9	300	11.8	771	CALD_CHICK	P12957 gallus galli
10	292.5	11.5	857	NFM_CHICK	P16053 gallus galli
11	291.5	11.5	700	TRDN_CANFA	P82179 canis famli
12	290	11.4	728	TRDN_HUMAN	Q13061 homo sapien
13	284	11.2	644	NFM_RABIT	P54938 oryctolagus
14	283	11.2	915	NFM_HUMAN	P07197 homo sapien
15	282.5	11.1	848	NFM_MOUSE	P08553 mus musculu
16	280.5	11.1	810	NFM_BOVIN	O77788 bos taurus
17	280	11.0	1220	IF2F_HUMAN	O60841 homo sapien
18	279	11.0	667	CYL1_BOVIN	P35662 bos taurus
19	278	11.0	488	CYL2_BOVIN	Q28092 bos taurus
20	274	10.8	2464	NAPB_MOUSE	P14873 mus musculu
21	273.5	10.8	845	NFM_RAT	P12839 rattus norv
22	270.5	10.7	598	CYL1_HUMAN	P35663 homo sapien
23	268.5	10.6	2459	NAPB_RAT	P15660 rattus norv
24	266	10.5	483	M6_STRPY	P08088 streptococc
25	265.5	10.5	492	M5_STRPY	P02977 streptococc
26	265.5	10.5	793	CALD_HUMAN	Q05682 homo sapien
27	260.5	10.3	1359	ATRX_CAEEL	O94760 caenorhabdi
28	260	10.2	1403	YDF3_SCHRO	Q10475 echinosach
29	259.5	10.2	2468	NAPB_HUMAN	P46821 homo sapien
30	258.5	10.2	2116	MIS2_DICDI	P08799 dictyosteli
31	257.5	10.1	678	GSRP_PLA6F	P13816 plasmodium
32	257.5	10.1	877	INCE_CHICK	P33352 gallus galli
33	257.5	10.1	4684	PLEI_HUMAN	Q15149 homo sapien

34	256.5	10.1	4687	PLEI_RAT	P30427 rattus norv
35	256	10.1	1790	USO1_YEAST	P25386 saccharomyc
36	255.5	10.1	564	M2_STRPY	P19401 streptococc
37	254	10.0	344	MSTI_DROHY	Q08695 drosophila
38	252.5	10.0	4473	PLEI_CRIGR	O9155 cricetus
39	252	9.9	704	NP14_RAT	P41737 rattus norv
40	251.5	9.9	1233	PF16_YEAST	P43597 saccharomyc
41	241	9.5	2492	ATRX_HUMAN	P46100 homo sapien
42	240.5	9.5	1500	SSP5_STRGN	P16952 streptococc
43	239.5	9.4	699	NP14_HUMAN	O14978 homo sapien
44	239	9.4	738	YK4_YEAST	P35732 saccharomyc
45	237.5	9.4	1962	MYSA_DROME	P05661 drosophila

ALIGNMENTS

RESULT 1

AC P12036, STANDARD, PRT, 1020 AA.
ID NFH_HUMAN
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurofilament triplet H protein (200 kDa neurofilament protein)
DE (Neurofilament heavy polypeptide) (NF-H).
GN NF-H OR NFH.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Carnivora; Homiidae; Homo.
NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88328981; PubMed=3138108;
RA Lees J.F., Sheldman P.S., Skuntz S.F., Carden M.J., Lazzarini R.A.;
RT "The structure and organization of the human heavy neurofilament subunit (NF-H) and the gene encoding it.";
RT EMBO J. 7:1947-1955(1988).
RL
CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M, AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT OBSERVED BY THE TWO SMALLER NF PROTEINS.
CC -1- PPM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE OF AXONAL CALIBER.
CC -1- PPM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC
EMBL, X15306; CA333366.1; -
EMBL, X15307; CA333366.1; JOINED.
EMBL, X15308; CA333366.1; JOINED.
EMBL, X15309; CA333366.1; JOINED.
PIR, S00979; OPHUH.
DR Genew, HGNC:7737; NEFH.
DR MIM, 162230.
DR GO, GO:0005883; C:neurofilament; NAS.
DR GO, GO:0007399; P:neurogenesis; NAS.
DR InterPro, IPR001664; IF.
DR Pfam, PF00038; filament; 1.
DR PROSITE, PS00226; IF; 1.

KW Intermediate filament; Coiled coil; Neurone; Phosphorylation.
 FT DOMAIN 1 100 HEAD.
 FT DOMAIN 101 413 ROD.
 FT DOMAIN 414 1020 TAIL.
 FT DOMAIN 101 132 COIL 1A.
 FT DOMAIN 133 145 LINKER 1.
 FT DOMAIN 146 244 COIL 1B.
 FT DOMAIN 245 266 LINKER 12.
 FT DOMAIN 267 288 COIL 2A.
 FT DOMAIN 289 292 LINKER 2.
 FT DOMAIN 293 413 COIL 2B.
 SQ SEQUENCE 1020 AA; 111780 MW; 1177C9D3DCFLD4 CRC64;

Query Match 14.3%; Score 363; DB 1; Length 1020;
 Best Local Similarity 25.9%; Pred. No. 5,5e-07;
 Matches 152; Conservative 79; Mismatches 239; Indels 118; Gaps 22;

QY 5 KSERVHYHSIRKFSVGASVVV-----ASLVMSGVVHATENEGATQVPTSSNRANESQAQ 59
 DB 439 KSEKI--KVEKSKETIYVEQTEFYVTEVEEKEKEKEEGKEEAGE 496
 QY 60 GEQPKLDSERDKA--RKEVEYK-KIVGESYAKSKRHTTVALVNLNNIKNEYLN 116
 DB 497 GSEETKSPAEBAASPEKEAKSPVKEAKSPAEAKSPAEVKSPEKASP--- 553
 QY 117 KIVESTSSEQOILMMESRSKVDENVKPEKSSSSSSSDSTREASDTAKPKPT-EP 175
 DB 554 --AKEBAKSPPEAKSPPEAKSPPEAKSPPEAKSPPEAKSPPEAKSPVKEE 611
 QY 176 GEKVAEAKKV-EEAKKADOKEDRRNYPTITTYKLELEIAESDVEYKALELVKVA 234
 DB 612 AKSPAEAKSPVKEAKSPAEVKSPEKASP-----KEAKSPPEKASP---KEA 660
 QY 235 NEPRDEQIKQAEAVEK-----QAEATRLKIKITD-REKEEAKRADAKEGKRG 288
 DB 661 KSPPEAKSPVNAEAKSPPEAKSPVNAEAKSPPEAKSPVKEAKSPPEAKSPPE 720
 QY 289 RAKGVPEELTPDCKENDAKSSDSSVGEETLPSLSLPEKVAEAKKVEAKKAKADQ 348
 DB 721 KAKSPVKEAKTPPEAKSPVKEAKS-----PEKASPPE-AKATLDVKSPEAKTPAKE 773
 QY 349 KEEDRRNYPTTYKLELEIAESD-----VEYKALEL-VKEAEKPEPNEE----- 394
 DB 774 ARSPADKPEPKAKSPVKEEVSPPEAKSPPEAKSPPEAKSPPEAKSPPEAKSP 833
 QY 395 KYQKAAVESKKAATRLKIKITDRKKAEEBAKKAABEDVKYK----- 440
 DB 834 KYKEPKKAAEEKAPATPKTEKDSKK--EEAPKKEAPKPVKEKPEPAVEKPKESKVE 891
 QY 441 ----PAEQOPAPAKAEKP-----APAKPENPAOPKAE 472
 DB 892 AKKEBAEKKKVPTPEKAPAKVAVKEDAKPKPEKTEVAKKEPDDAKAEKSPPEAKKBA- 950
 QY 473 KPADOQAEEDYARSEEEYNRLTQOQPKTEKPA-----QSTP 511
 DB 951 APEKDTPEKAKKPEK-----PKTEKAKEDDKTISKERSKP 989

RESULT 2
 BAG_STRAG STANDARD; PRT: 1164 AA.
 AC P2751;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE IGA FC receptor precursor (Beta antigen) (B antigen).
 GN BAG.
 OS Streptococcus agalactiae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OC NCBI_TaxID=1311;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 38-48.

RC STRAIN=IA239;
 RX MEDLINE=91312121; PubMed=1857207;
 RA Uerlstrom P.G., Chhatwal G.S., Timms K.N.;
 RT "The IGA-binding beta antigen of the c protein complex of Group B
 RT streptococci: sequence determination of its gene and detection of two
 RT binding regions.";
 RL Mol. Microbiol. 5:843-849(1991).
 RN [2]
 RP IDENTIFICATION OF IG-LIKE DOMAIN.
 RX MEDLINE=97035265; PubMed=8880921;
 RA Bateman A., Eddy S.R., Chochia C.;
 RT "Members of the immunoglobulin superfamily in bacteria.";
 RL Protein Sci. 5:1939-1942(1996).
 CC -I- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
 CC an amide bond (potential).
 CC -I- SIMILARITY: Contains 1 immunoglobulin-like domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use. By non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X59771; CAA42442.1; .
 CC PIR: S15330; PCSOAG.
 CC InterPro: IPR004829; Gaurface_antigen.
 CC InterPro: IPR005877; Gpos_YsIRK.
 CC InterPro: IPR001899; Gram_pos_anchor.
 CC InterPro: IPR003599; IG.
 CC InterPro: IPR006192; LPXTG.
 CC Pfam: PF00746; Gram_pos_anchor; 1.
 CC Pfam: PF00662; RICH; 1.
 CC Pfam: PF04650; YsIRK_signal; 1.
 CC ProDom: PD153432; Gaurface_antigen; 1.
 CC SMART: SM00409; IG; 1.
 CC TIGRFAMs: TIGR01167; LPXTG_anchor; 1.
 CC TIGRFAMs: TIGR01168; YsIRK_signal; 1.
 CC DR PROSITE: PS50847; GRAM_POS_ANCHORING; 1.
 CC DR PROSITE: PS50835; IG_LIKE; FALSE_NEG.
 CC KW Cell wall; Peptidoglycan-anchor; Receptor; Repeat; signal;
 KW Immunoglobulin domain.
 FT CHAIN 1 37
 FT PROPEP 136 1135
 FT DOMAIN 434 534
 FT DOMAIN 199 438
 FT DOMAIN 439 826
 FT DOMAIN 827 945
 FT SITE 1132 1136
 FT MOD RES 1135 1135
 SQ SEQUENCE 1164 AA; 131051 MW; 65DE94AF720A5474 CRC64;
 Query Match 13.9%; Score 351.5; DB 1; Length 1164;
 Best Local Similarity 28.2%; Pred. No. 1.6e-06;
 Matches 143; Conservative 82; Mismatches 177; Indels 105; Gaps 21;
 QY 1 MFASKSERKHYHSIRKFSVG-ASVVAASLVMSGVVHATENEGATQVPTSSNRA---NES 55
 DB 1 MFKNYERKMYHSIRKFSVGASVAASLVMSGVVHATENEGATQVPTSSNRA---NES 60
 QY 56 QAEQGEQPKLDSERDKA-----KVEEYVYKIVGESYAKSKRHTTVALVNLNN 109
 DB 61 QTDGNNSSSELETTTWEIPTTDIKAAVEVEKTAETSATDQKKEK---QLOQWKNN 117
 QY 110 IKNEYLVKIVESTESQOILMMESRSKVDENVKPEKSSSSSSSDSTREASDTAKP 169
 DB 118 LKNDVDTILSHQKN-----EFKTKID-----TNDSDALLE----- 152
 QY 170 NKPTPEKGV-AEAKKVEEAKKADOKEDRRNYPTITTYKLELEIAESDVEV--KKA 226
 DB 153 NQFNETRLHLIKQHEEVEKQKAKQK-----TLKQSDTVADSNIDKELNHQSQ 204

QY 227 LELVVKANEPBDEQIKQAEVAVESKQAEATRLKIKITDREBEAEKARRADA-----K 281
 Db 205 VE-----KMAEQKGTINEDKOSMLKKIKIDIRKQA-QQADKEDAEVAKRE 248
 QY 282 EOGKPKGAKGVPELATPPKENDAKSSDSVGEETLPPSPSLPEKKVAEAKKVEEA 341
 Db 249 ELGKLFSSTKAGLDQEIQEHVYKETSSEBNTQKVDHVAHSLQNLQKSLBELDKAT--- 305
 QY 342 KKKAEQOEEDRRNPTVTYTLLEL-IAESDVEVKA---ELELVKEAEKPEPNEE- 394
 Db 306 ---TNEQATQVKNQFLENAQKLEIQPLIKETNVLKYKAMESLEQVKEKLEK--HNSKAN 360
 QY 395 -----KYQAKAEVSKKAEATRLKIKITDRKAEAEKAKR---AAEDKVKKEKPAOP 445
 Db 361 LEDVAKSEKIVREYEGKLNOSKLP----KQLEBAHSLKQVVEDPFRKFKTTSQV 416
 QY 446 QPAPAPKAEKPAKPPENPAOPKAE 472
 Db 417 TPKKRVKADLAA-----NENNOQKIE 437
 RESULT 3
 MST2 DROHY STANDARD; PRT; 1391 AA.
 AC Q08696;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Axoneme-associated protein mst101(2).
 GN Mst101(2).
 OS Drosophila hydei (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Peiryoeta;
 OC Ephydroidea; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC NCI_TaxID=7224;
 RN [1]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RX MEDLINE=9504538; PubMed=7957199;
 RA Neesen J., Padmanabhan S., Buemann H.;
 RT Tandemly arranged repeats of a novel highly charged 16-amino-acid
 RT motif representing the major component of the sperm-tail-specific
 RT axoneme-associated protein family Dmst101 form extended
 RT alpha-helical rods within the extremely elongated spermatozoa of
 RT Drosophila hydei.";
 RL Eur. J. Biochem. 225:1089-1095(1994).
 CC -1- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- TISSUE SPECIFICITY: TESTIS. PRIMARY SPERMATOCTES AND EARLY
 CC SPERMATIDS.
 CC -1- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.
 CC -1- POLYMORPHISM: LENGTH POLYMORPHISMS EXIST BETWEEN DIFFERENT
 CC STRAINS, MOST LIKELY CAUSED BY LENGTH VARIATIONS WITHIN THE TANDEM
 CC REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, X73481, CAA51876.1, -.
 DR PIR; S51364; S51364.
 DR FlyBase; FBgn0020733; Dhyd\mst101(2).
 KM Sperm; Repeat; Multigene family; Polymorphism.
 FT DOMAIN 332 1268 [KR]-K-X-C-X-X-X-A-X-A-X-X-X-X-X-X-X-E.
 FT SEQUENCE 1391 AA; 159000 MW; 1B2A368F30F48878 CRC64;
 SQ
 Query Match 12.9%; Score 328; DB 1; Length 1391;
 Best Local Similarity 27.0%; Pred. No. 1.3e-05;

Matches 140; Conservative 79; Mismatches 220; Indels 80; Gaps 15;
 QY 38 ENEGATVPTISSNANE-SQAEQEGOPRKUDSER---DKAEKVEEYKIVGSSYAKST 93
 Db 363 EKKAKCELAKKKKKAEDEKKEEAEKKEAEKKEKKEKKEKKEKKEKKEKKEKKEKKE 422
 QY 94 KKRHTITVALVLENNINIKNEYLANKIVESTSESQOILIMESRSYVDEAVSFEDSSSSS 153
 Db 423 KE-----AAERKCEELAKNIKKAEEK--KCEAKKEKKEAEKKEKCEELAKK 469
 QY 154 SSSSTKPEASDTAPKPKPTPEGKVAEAKKVEAKKAKQOEEDRRNPTVTYTLLE 212
 Db 470 IKKAEKKKCEETAKKGEVAKKCEELAKKIKKAEIKKKCKKLAKKEREKKEKCEK 529
 QY 213 -----LEIAESD-----VEYKAELELVKIANPDEQIKQA-----PAEV 250
 Db 530 AAKKKEAEKKEKCEKAKKREAEKKEKCEKAKKRR--EAAEKKEKCEKAKKEKAE 587
 QY 251 ESKQAEATRLKIKITDREBEAEKARRADAKQKPGRAK-----RGVPELATPDK 303
 Db 588 KKKCEEAKEKEVAERKKECEELAKKIKKAEKKEKCEAAKKEKKEAEKKEKCEEA 647
 QY 304 KENDAKSSDSVGEETLPPSLKPE-----KVAEAEKVEEA--KKAEQOK 349
 Db 648 KAAEKKCKKLAKKEREKTAERKKEKCEKAKKREAEKKEKCEAAKKEKKEKCEEA 707
 QY 350 EEDRRNPTVTYTLLELIEASDVEVKAELVKEAEKPEPNEEKYQAAEVEKKA 409
 Db 708 KKEEAERKKECEELAKKIK-----KAAEKKCKKLAKKAKKAEKKNLKKGNKGA 760
 QY 410 ATRLKIKITDRKAEAEKRAAEEDVKEKEKPAOPAPAPKAEKPAKPPENPAOP 469
 Db 761 -----LKEKKCKRELAKKAAEKKCKEAKKKEKAEKKEKCEKAKKKEE--AEKK 811
 QY 470 KAEKPADQAEEDYARSEE-EYNRLTQOQPPKTEKPAQ 507
 Db 812 KCEKTAKKRKEAEKKEKCEKAKKREAEKKEKCEKTA 850
 RESULT 4
 UN89 CAEEL STANDARD; PRT; 6632 AA.
 AC 001761; Q17362;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).
 GN UNC-89 OR CO9D1.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodertinae; Caenorhabditis.
 OC NCI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
 RC STRAIN=Bristol N2;
 RX MEDLINE=86180278; PubMed=8603915;
 RA Benian G.M., Tinley T.L., Tang X., Borodovsky M.;
 RT "The Caenorhabditis elegans gene unc-89, required for muscle M-line
 RT assembly, encodes a giant modular protein composed of Ig and signal
 RT transduction domains";
 RL J. Cell Biol. 132:835-846(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RA Du Z., Le T.T., Wilson R.;
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP REVISIONS.
 RA Waterston R.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Structural component of the muscle M-line. Myofibrilment
 CC lattice assembly begins with positional cues laid down in the
 CC basement membrane and muscle cell membrane. UNC-89 responds to

Db 1485 KKEKSPATNIVEVSEETITETITMTMTHSE-----ESRTSVKKEKTEPEYDEK 1538
 Qy 151 SSSSSDSTKEASDTAK---PNKPTGEKVAAE-----AKKVEEAKKADQKEDRR 201
 Db 1539 PKSPKCKDCKSEKITEIKSPVKKESPEKVEKPSPTKKEKSPKPSPTKKESENV 1598
 Qy 202 NYPTTYKTLELEAESDVEVKAELELVKVAKEPRDEQIKQAEAEVSKQAEATRLK 261
 Db 1599 KSPFKKEKSEKPEKVE-----ELKSPKKESEKADK--KPKSPKKEKSEKSAEDV 1649
 Qy 262 KIKTDREBAE--EAKRAADAKKEGKP-----KRAKRGVPGELATPDKEN 306
 Db 1650 KSPFKKEKSEKPEKVEKPSPTKKESSPTKCTDVEKSPKKEKSPQVTEKPAKSPKKE- 1708
 Qy 307 DAKSDSDSVGEETLPSPSLKPEKVAEAEKKEVKAQKEDRNRNTYNTYKTLEL 366
 Db 1709 --KSPKESVVEKSPKKEKSPK--AEKPSPTKKEKSPKSAAEVKSPTKKEKSPK 1764
 Qy 367 EIAESDVEVKAELELVK---EKAEPNNEKVKQAKAEVSKKAEATRLKIKTDKKA 423
 Db 1765 SAEKPKSPKTKESSPVMADDEVKSPKKEK---SPKVEKPAKSPKKE--KTPKSA 1819
 Qy 424 EBEAK-----RKAEBDKVK--EKPAEQPAPAPAEKAPAPAKPENPAEQ 468
 Db 1820 AEBKSPKTKKEKSPSPKTKGTGDESKKESPEKPEKPK--SPTPK-KSPGSPK-KKSKS 1876
 Qy 469 PKAEKPADQQAED 482
 Db 1877 PEAKPPAPKLTTRD 1890

RESULT 5

NFH_RAT ID_NFH_RAT STANDARD; PRT; 831 AA.

AC P16884; Q63368;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Neurofilament triplet H protein (200 kDa neurofilament protein)
 DE (Neurofilament heavy polypeptide) (NF-H) (Fragment).
 GN NFH OR NFH.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus;
 OC NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=89065087; PubMed=3143606;
 RA Breen K.C., Robinson P.A., Wion D., Anderson B.H.;
 RT "Partial sequence of the rat heavy neurofilament polypeptide (NF-H).
 RT Identification of putative phosphorylation sites.";
 RL FEBS Lett. 241:213-218(1988).
 RN [2]
 RP SEQUENCE OF 37-831 FROM N.A.
 RX MEDLINE=88309090; PubMed=2457365;
 RA Dautigny A., Pham-Dinh D., Kousseil C., Felix J.M., Nussbaum J.L.,
 RA Jolles P.;
 RT "The large neurofilament subunit (NF-H) of the rat: cDNA cloning and
 RT in situ detection.";
 RL Biochem. Biophys. Res. Commun. 154:1099-1106(1988).
 RN [3]
 RP SEQUENCE OF 1-89 AND 243-313 FROM N.A.
 RX MEDLINE=87080760; PubMed=2878828;
 RA Robinson P.A., Wion D., Anderson B.H.;
 RT "Isolation of a cDNA for the rat heavy neurofilament polypeptide
 RT (NF-H).";
 RL FEBS Lett. 209:203-205(1986).
 RN [4]
 RP SEQUENCE OF 318-831 FROM N.A.
 RX MEDLINE=9914647; PubMed=2928342;
 RA Liebung I., Spilner N., Snyder S., Anderson J., Goldhaber D.,
 RA Smulowitz M., Carroll Z., Emanuel B.S., Breitner J., Rubin L.;

RT "Cloning of a cDNA encoding the rat high molecular weight
 RT neurofilament peptide (NF-H): developmental and tissue expression in
 RT the rat, and mapping of its human homologue to chromosomes 1 and
 RT 22.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
 CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.
 CC -1- PPM: THERE ARE A NUMBER OF REPEATS OF THE TRIPLETT K-S-P, NFH IS
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
 CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
 CC OF AXONAL CALIBER.
 CC -1- PPM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 783
 CC ONWARD AND IS LONGER DUE TO A FRAMESHIFT.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by, and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: M37227; AAA41693.1; ALT_FRAME.
 DR EMBL: X13804; CAA32038.1; ALT_FRAME.
 DR EMBL: M21964; AAA41695.1; --
 DR EMBL: J04517; AAA41692.1; --
 DR InterPro: IPR001664; IF.
 DR Pfam: PF00038; Filament; 1.
 DR PROSITE: PS00226; IF; 1.
 KW Intermediate filament; Coiled coil; Neurone; Phosphorylation; Repeat.

FT NON_TER 1
 FT DOMAIN 1 1
 FT 276 641
 FT 164 164
 FT 185 185
 FT 193 193
 FT 199 199
 FT 346 346
 FT 373 373
 FT 482 482
 FT 485 485
 FT 570 571
 FT 591 591
 FT 727 727
 FT 757 757
 FT 769 769
 FT 775 775
 SQ SEQUENCE 831 AA; 89486 MW; 1B0973C3F13EF68 CRC64;

Query Match 12.5%; Score 316; DB 1; Length 831;
 Best Local Similarity 25.3%; Pred. No. 2.3e-05;
 Matches 150; Conservative 82; Mismatches 213; Indels 148; Gaps 23;

Qy 40 EGATQVPTSNRAN-----EQAQGEQPKLDSRDARKEVEE 79
 Db 191 EGIPLKPSWSTHIKVKSEKIKVSEKETEIVVEQTEIQTVEETEDDEAQQEE 250
 Qy 80 YVKKIGSEYAKSTKGRHTITVALVNLNKKVEYLNKIVSETSQLOQLIMMESRKYD 139
 Db 251 EHEBEGGBAATSPPAEBAASPEKTKSPVVE-----AKSPAEEKSP--AEAKSPA- 301
 Qy 140 EAVSKTEKDSSSSSSSDSSSTK-----PEASDTAKPKPKTE-----PGE----- 177
 Db 302 EAKSPAEEVSPVAKSPAEEVSPVAKSPAEEVSPVAKSPAEEVSPVAKSPAEEVSPVAKSPA 361
 Qy 178 -----KVAEAKKVE-----EAKKADQKEDBRNRNPTITTYKTLELEI-- 215

```

Dd      362  EAKSPAEBVKSPAEAKSPAEBVKSPEGEAKSPAEBKSPAEBVKSPPATVKSPEVEAKSPA 4211
Qy      216  -----AESDVEVKKALETIVYKANEPRDEOKIKQAEAEVESKO--AEATRLK 261
Dd      422  EYKSPVTVYKSPAEBKSPVBEVKSPPASVKSSEAKSPGAKSPAEBKSPVVAKSPAEBKSPA 481
Qy      262  K1KTDR- -EAEBEAKRRADAKEGKPKGAKRGVPELATTPDKEN-----DAKS 310
Dd      482  GAKPPAEAKSPAEBKSPAEBKSPAEBKSPVBEVKSPEAKSPVKGAKSLAEAKS 541
Qy      311  SD---SSVGEETLPSPSLK-----PEKKVAEAKKV-----EAAKKKA 345
Dd      542  PEKAKSPVKEELIKPAEBVKSPEKAKSPMKKEAKSPKAKTLDVKSPBEAKPPAKEBAKRA 601
Qy      346  EDQKEEDRRNYNTNYTKTLELETAESDVEYKKALELVKEBAKEPPNBEKVQAKAEVES 405
Dd      602  DIRSPEQVKSPAKEBAKSPEKE-----ETRTEKVAPKKEEVKSPEVEVAKPEPKPVEE 655
Qy      406  KKAETIRLEK1TDRK-KAEER- -KKAABEDKVKEKPAEQOPAPAPKA- -KPPAPAKP 462
Dd      656  EKTPTPTKTEVESKDEAPKREOKPAEKEBELTEKPKXDSPEAKKEEAKKEKKAAPEE 715
Qy      463  ENPAP-----QPAKEKPADQAEEDYARSRSEETNRLTQOQPPTEKRAOP 508
Dd      716  ETAPKLVGKEAKPK- -EKAEDAKAKP- -SKPSKE-----KPKKEEVPAP 759

```

RESULT 6	ID	STANDARD	PRT:	705 AA.
TRDN_RABIT				
AC	Q28637	Q28637; Q28643;		
DT	30-MAY-2000	(Rel. 39, Created)		
DT	30-MAY-2000	(Rel. 39, Last sequence update)		
DT	15-SEP-2003	(Rel. 42, Last annotation update)		
DE	Triadin.			
GN	TRDN.			
OS	Oryctolagus cuniculus (Rabbit).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.			
OX	NCBI_Taxid=9986;			
RA	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORMS SKELETAL).			
RC	STRAIN=New Zealand white; TISSUE=Skeletal muscle;			
RX	MEDLINE=93286104; Pubmed=7685347;			
RA	Knudson C.M., Steng K.R., Moomaw C.R., Slaughter C.A., Campbell K.P.,			
RT	"Primary structure and topological analysis of a skeletal muscle-			
RT	specific junctional sarcoplasmic reticulum glycoprotein (triadin).";			
RL	J. Biol. Chem. 268:12646-12654 (1993).			
RA	[2]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORMS SKELETAL).			
RC	TISSUE=Skeletal muscle;			
RX	MEDLINE=94298946; Pubmed=8026576;			
RA	Peng M., Fan H., Kirley T.L., Caswell A.H., Schwartz A.;			
RT	"Structural diversity of triadin in skeletal muscle and evidence of			
RL	its existence in heart.";			
RL	FEBS Lett. 348:17-20 (1994).			
RA	[3]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORMS CARDIAC).			
RC	TISSUE=Heart muscle;			
RX	MEDLINE=96132942; Pubmed=8550602;			
RA	Guo W., Jorgensen A.O., Jones L.R., Campbell K.P.;			
RT	"Biochemical characterization and molecular cloning of cardiac			
RL	triadin.";			
RL	J. Biol. Chem. 271:458-465 (1996).			
RA	[4]			
RP	CARBOHYDRATE-LINKAGE SITES, AND INTERCHAIN DISULFIDE BONDS.			
RX	MEDLINE=96066664; Pubmed=7578102;			
RA	Fan H., Brandt N.R., Caswell A.H.;			
RT	"Disulfide bonds, N-glycosylation and transmembrane topology of			
RL	skeletal muscle triadin.";			
RL	Biochemistry 34:14902-14908 (1995).			
CC	-1- FUNCTION: MAY BE INVOLVED IN ANCHORING CALSEQUESTRIN TO THE			

CC	JUNCTIONAL SARCOPLASMIC RETICULUM AND ALLOWING ITS FUNCTIONAL COUPLING WITH THE RYANODINE RECEPTOR.
CC	-I SUBUNIT: Homooligomer of variable subunit number; disulfide-linked.
CC	-I SUBCELLULAR LOCATION: Type II membrane protein. Sarcoplasmic reticulum.
CC	-I ALTERNATIVE PRODUCTS:
CC	Event=Alternative splicing; Named isoforms=6;
CC	Comment=Additional isoforms seem to exist;
CC	Name=Skeletal 1; Synonyms=ST1;
CC	Isoid=Q28820-1; Sequence=Displayed;
CC	Name=Cardiac 1; Synonyms=CT1;
CC	Isoid=Q28820-2; Sequence=VSP_004458, VSP_004460;
CC	Name=Cardiac 2; Synonyms=CT2;
CC	Isoid=Q28820-3; Sequence=VSP_004459, VSP_004461;
CC	Name=Cardiac 3; Synonyms=CT3;
CC	Isoid=Q28820-4; Sequence=VSP_004466;
CC	Name=Skeletal 2; Synonyms=ST2;
CC	Isoid=Q28820-5; Sequence=VSP_004462, VSP_004463, VSP_004464,
CC	VSP_004465;
CC	Name=Skeletal 3; Synonyms=ST3;
CC	Isoid=Q28820-6; Sequence=VSP_004464, VSP_004465;
CC	-I TISSUE SPECIFICITY: SKELETAL AND CARDIAC MUSCLE.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collabora-
CC	tion between the Swiss Institute of Bioinformatics and the EMBL outstati-
CC	c at the European Bioinformatics Institute. There are no restrictions on
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commer-
CC	entiaes requires a license agreement (see http://www.isb-sib.ch/annou-
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; U31540; AAC48496.1; -
DR	EMBL; L10065; AAA31488.1; -
DR	EMBL; U31555; AAC48497.1; -
DR	EMBL; U34201; AAC48498.1; -
DR	PIR; A45390; A45390.
KM	Transmembrane; Sarcolemmal reticulum; Glycoprotein;
KW	Alternative splicing.
FT	INIT MET 0
FT	DOMAIN 1 46
FT	TRANSMEM 47 67
FT	DONAIN 68 705
FT	CARBOHYD 74 74
FT	CARBOHYD 624 624
FT	VASPLIC 264 285
FT	
FT	
FT	VASPLIC 264 307
FT	
FT	
FT	VASPLIC 286 705
FT	
FT	VASPLIC 308 705
FT	
FT	VASPLIC 415 415
FT	
FT	VASPLIC 416 424
FT	
FT	VASPLIC 584 584
FT	
FT	VASPLIC 585 591
FT	
FT	VASPLIC 645 705
FT	
FT	
SEQ	SEQUENCE 705 AA; 79003 MM; 13AFIDB4475A1361 CRC64;

12.4%; Score 315.5; DB 1; Length 705;

Query Match

12.4

Score 31

1; Length 705;

	Best Local Similarity 25.0%: Pred. No. 2,1e-05: Matches 135; Conservative 83; Mismatches 195; Indels 127; Gaps 21
Qy	38 NEGATQVPTSSNRANESQAEQGEOPKLDISERDARKVEVEYVKIIGESYAKSTKRH 97
Db	155 ERKIPTVVHKEKEKEKEKVEKEKEPEKKAATHKEKLEKEKEPETVTVKEKKANTKEK- 213
Qy	98 TITVALNELNNIKVEYLKIVESTSESOQLIMMESKSKUDAVSKPEKSSSSSS- 155
Db	214 -IEETKKEVGVQKEKVKQTAKAKEVQKTPRKPEKESKETAAVSKOEKQVAFCHYM 272
Qy	156 ----DSSTKEASDTAKPNKTE-----PGEKVAE---AKKVEAKKAKOK 196
Db	273 IDIFVHGDLKGEQSPALPPSPETBOASRPPTPALPTPEEKEGEKKAEKKTVTETKKAEK 332
Qy	197 EEDRRNPTITYKLTLELIAESDVEVKKAELVLYKANPEDEOKIKQAEAEVSKQAE 256
Db	333 EDAKK-----SEKETDIDMK-----KEPKSPDTPTGTYKVTQA- 370
Qy	257 ATRLKIKITDREBEAEBAKR-RADAKEGOKPKRAKGVPGELATPDKKENDAKSSDVS 315
Db	371 ----TKDKEKEDSKKAKRAPEEQPKGKQKEKKEHEPAKSTYKKEHA----- 415
Qy	316 GEETLPSPLPREKVVAAEKVBEAK-KKAEQKEEERNRYNPTITYLTLELIAE- 370
Db	416 ----PSEKQAKIKERKEEVSASATKAVPAKKEK-----TITVQETRKERKG 462
Qy	371 ----SDVEY-KKAELEL---VKEAKEPRNEEKVKQAKAEVSKKAEATYLE----- 414
Db	463 KISSVLMDKELTKEVEVAVPASPENKSGSETKDEKT-SKEPQIKKEEKQGVKPKRP 520
Qy	415 --KIKTRDKKAAEBEAK-RKAAEDKVKEKPAEQOPARAPAKREKPAKPEKPNPAQOP- 469
Db	521 QPQIKKEEKPRQDINKPEKTLHGKPEKVKLVQVAVATTEKHVVKPKPAKKEHCKEPPS 580
Qy	470 -KAEKPAD-----QQAEEDY---ARSEEEYNRLTQQQPKTEKPAOPS 509
Db	581 IKTKPKSTSGKMPVETESGKKIKLEKSEKIKVPAKRRESHOLQNTV----KAEKPARGS 635
RESULT 7	
M24_STREPY STANDARD; PRT; 539 AA.	
ID_M24_STREPY	
AC P12379;	
DT 01-OCT-1989 (Rel. 12, Created)	
DT 01-OCT-1989 (Rel. 12, Last sequence update)	
DT 28-FEB-2003 (Rel. 41, Last annotation update)	
DE M protein, serotype 24 precursor.	
GN EMM24.	
OS Streptococcus pyogenes.	
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;	
OX Streptococcus.	
NCBI_TaxID=1314;	
RN [1]	
RA SEQUENCE FROM N.A.	
RC STRAIN=Vaughn / Serotype M24;	
RX MEDLINE=8811516; PubMed=3276665;	
RA Mouw A.R., Beachey E.H., Burdett V.;	
RT "Molecular evolution of streptococcal M protein: cloning and	
nucleotide sequence of the type 24 M protein gene and relation to	
other genes of Streptococcus pyogenes.";	
RL J. Bacteriol. 170:676-684(1988).	
CC -1- FUNCTION: THIS PROTEIN IS ONE OF THE DIFFERENT ANTIGENIC SEROTYPES	
OF PROTEIN M. PROTEIN M IS CLOSELY ASSOCIATED WITH VIRULENCE OF	
THE BACTERIUM AND CAN RENDER THE ORGANISM RESISTANT TO	
PHAGOCYTOSIS.	
CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by	
an amide bond (potential).	
CC -1- SIMILARITY: TO OTHER M PROTEINS.	
CC	
CC This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC the European Bioinformatics Institute. There are no restrictions on its	

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M19031; AAA26874.1; -.
DR PIR: A28549; A28549.
DR InterPro: IPR005877; Gpos_Y5IRK.
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR006192; LPTXG.
DR InterPro: IPR003345; M_repeat.
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR Pfam: PF02370; M_11.
DR Pfam: PF04650; Y5IRK_signal; 1.
DR PRINTS: PR00015; GPOSANCHOR.
DR TIGRfams: TIGR01167; LPTXG_anchor; 1.
DR TIGRfams: TIGR01168; Y5IRK_signal; 1.
DR PROSITE: PS50847; GRAM_POS_ANCHORING; 1.
KW virulence; Phagocytosis; Cell wall; Peptidoglycan-anchor; Repeat;
KW Antigen; Collid coll; Signal.
FT SIGNAL 1 42
FT CHAIN 1 42
FT PROPEP 509 539 M PROTEIN, SEROTYPE 24.
FT DOMAIN 118 301 REMOVED BY SORTASE (POTENTIAL).
FT REPEAT 118 152 5.3 x 35 AA TANDEN REPEATS, A-TYPE.
FT REPEAT 153 187 A-1.
FT REPEAT 188 222 A-2 (1 ALTERATION).
FT REPEAT 223 257 A-3 (2 ALTERATIONS).
FT REPEAT 258 292 A-4 (1 ALTERATION).
FT REPEAT 293 301 A-5 (2 ALTERATIONS).
FT DOMAIN 311 301 A-6 (INCOMPLETE).
FT REPEAT 311 355 2.7 x 35 AA TANDEN REPEATS, B-TYPE.
FT REPEAT 356 380 B-1.
FT REPEAT 381 405 B-2.
FT DOMAIN 468 504 GLY/PRO-RICH (CELL WALL-SPANNING).
FT SITE 505 509 LPTXG SORTING SIGNAL (POTENTIAL).
FT MOD_RS 508 508 AMPDE-LINKED TO CELL WALL (POTENTIAL).
SQ SEQUENCE 539 AA; 58804 MW; B03EDPF3AC1BE69C7 CRC64;

Query Match 12.3%; Score 311; DB 1; Length 539;
Best Local Similarity 27.5%; Pred. No.2.5e-05;
Matches 157; Conservative 90; Mismatches 200; Indels 124; Gaps 27;

QY 4 SKSERKHYVSIIRKFSVGS-ASVVAASLWGSVVAHTENEGATQVPTSSNRANESQAEQ 62
DB 2 TKNTNTNHYSLRLKTKCTASVAVALTYLGAGLVNNTNE -VSAVATRS-----Q 48
QY 63 PKKIDSRDQARK-EVEEYVKYKIGVESYAKSTKKRHTTYALVNLNINKEVYANKIVES 121
DB 49 TDTLEKQVEADRFELIENNTLKLKNSDLSFNFNKALKOHNDLFEELSNAK-EKLRKDKDS 107
QY 122 TTSSEQQLIMMESR-----SKYDEVSK---FEKQSSSSSSSDSSTRE- 162
DB 108 LSEKASKIQLELEAKADLEKALBEGAMNFTADSASIKITLEBAEKALARKADLEKALEGA 167
QY 163 ----ASPTAKPNKPTPEGKVAEAKKKVEEAKKAKOKE-EDBRNYPY---ITYKITLELE 214
DB 168 MNFTADSAK-----IKTLEBAEKALBARQAELEKALEGAMNFTADSASIKITLEAE 219
QY 215 IASDVEVKKAELELVKVKANE-----PRDEOKIKOAEAE---VESKOAEPATRL----- 260
DB 220 --KAALLARKADLE---KALEGAMNFTADSASIKITLEBAEKALBARQAELEKALEGAM 273
QY 261 -----KKTKT--DREAEFEAKRRADAKKQCKPRKGRKGVGEL-----ATP 301
DB 274 NFSTADSAKIKTLEAEKALAEAR-KADLEHQSUVLANNRQSLRRDDASREAKKOLEAEH 332
QY 302 DKENDAKSSDSSVGEEETLPSPLKCEKVAEAE-KVEEBAKKAEQOKEEDRRNYPNT 360
DB 333 QKLEQNKISSEARQSLRRDLDSREAKKOLEHOKLEQNKISSEARQSLRRDDASR 392
QY 361 YKTLLEIASDVEVKKAELELVKEBAKPRNEEKVQKQAEVYSK-KADATRL-EXIKT 418
```

```

Db      333 EAKKQVEKALBEANSLKALALELTKNLEEE--SKULTEKKAQLQKLEKFAKALKEKL-- 448
Qy      419 DRKKAEEBAKRYKAAEDVKYKPKADQPODAPAPKAEKPAKPKPENPAEQPAEKP-ADQ 477
Db      449 -AKQAEELAKIRAGK-----ASDSQTPDPAKPGNKAVPGKGAPOAGT 489
Qy      478 QAEDYVARSSEENRLTQQQPKTEKPAQ 508
Db      490 KPNQKAPKPK-----TKRQLPSTGETAMP 514

RESULT 8
NFH_MOUSE
ID_NFH_MOUSE STANDARD; PRT; 1087 AA.
AC P19246; Q61959;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neurofilament triplet H protein (200 kDa neurofilament protein)
DE (Neurofilament heavy polypeptide) (NF-H).
DE NEFH OR NFH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=89121513; PubMed=3220257;
RA Julien J.-P., Cole F., Beaudet L., Sidky M., Flavell D., Grosveid F.,
RA Mushynski W., and structure of the mouse gene coding for the largest
RT "Neurofilament subunit."
RL neurofilament subunit."
RN Gene 68:307-314(1988).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=89089138; PubMed=3145094;
RA Sheidman P.S., Carden M.J., Lees J.F., Lazzarini R.A.;
RT "The structure of the largest murine neurofilament protein (NF-H) as
RT revealed by cDNA and genomic sequences."
RL Brain Res. 464:217-231(1988).
RN (3)
RP SEQUENCE FROM N.A.
RC STRAIN=Swiss Webster; TISSUE=Brain;
RA Carden M.J.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
CC SUBSERVED BY THE TWO SMALLER NF PROTEINS.
CC -1- PUT: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC OF AXONAL CALIBER.
CC -1- PUT: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 534
CC TO 716 AND IS SHORTER DUE TO FRAMESHIFTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; M24496; AAA39813.1;
EMBL; M23349; AAA39813.1; JOINED.
EMBL; M24494; AAA39813.1; JOINED.

```

Query	Match	12.2%	Score 310	DB 1	Length 1087
Best Local Similarity	24.8%	Pred. No. 4.9e-05			
Matches 150	Conservative 75	Mismatches 23	Indels 144	Gaps 25	
DB	18	SVGASVWVASLWMSVWVHATEN-----EGATQV--PTSSNRANESQAEQGEQPKLDSER	70		
QY	426	SISIHNIKVSSEMIKVKVEKSEKEVYVIGQVEELRVGCVTEEDDEKKAQGEGBAEGBE	485		
DB	71	DKAREVEEYKVIQGE--SYAKSTKKRHITVALVNELNNIKQENYLNKIVESTSESOLQ	128		
QY	486	EKEEBELLAATSPPAEEMASPEKETKSR-----VKGE-----AKSPGEAK--	525		
DB	129	ILWMSRKRVDENAVSKPEKSSSSSSSSSTKRP--EASDTAKPKMPTREGKVAEAKKKV	186		
QY	526	-----SPEEAKSPAEAKSPGEAKSPGEAKSPGEAKSPAEKSPAEKSPAEKSPAEKSPA	576		
DB	187	E-----EAKKAKDOKEEDRRNVPITTYKLELEI-----AESDVEKKAE	226		
DB	577	EPKSPATVYKSPGEAKSPGEAKSPAEAKSPAEAKSPAEAKSPAEAKSPAEAKSPA	636		
QY	227	LELVKVKANEDRDQIKQAEAEVESHQ-AEATRLKIKITDRE-----EAEIEAKRA	278		
DB	637	TVKSPGEAKSPGEAKSPAEAKSPAEAKSPAEAKSPAEAKSPAEAKSPAEAKSPA	696		
QY	279	DAKQKGRPKGAAGVPEGLATPD--KKENDAKS--SSSSGVEGLTSPSLKPEKVAEA	334		
DB	697	EVKSPAEAKSPAEAKSPGEAKSPAEAKSPAEAKSPAEAKSPAEAKSPAEAKSPA	756		
QY	335	EKK-----VEEAKKKAADQKEEDRRNVPN-----TYKLELEIAESDV--	373		
DB	757	EAKSPIEVKSPEKAKTPVKEGAKSPAEAKSPGEAKSPYKEDIKPPAEAKSPKAEKSPYKE	816		
QY	374	-----EYKKALELEI--VKEEAKPRNEEKVKQKAEVESHQAEATRLKIKITDRK	421		
DB	817	GAKPPEKAKPLDVSPEAQTPVQGEALVPTDIRPEQVKSAPK-EKAKSPKEAKTKSEK	875		
QY	422	KA--EEAKKRAAEEDVKYKEK--AEQDQAPAPKA-----EKPAKPKP-----E	463		
DB	876	VAPKKEEVKSPKEEVAKKEPPKVEEKKTLPTPKTEAKESKQDQAPAEKAPKVEEKK	935		

QY 464 NPAOPK---AEKPADQQAEDYARRSEEE-----VNRLTQQQPPKTE 503
DB 936 TTEPKOSTABAKKEHAGEKKAAVASEETPAKLGCKEAKPKKTEKTTTTEADDTAK 995
QY 504 KPAQPS 509
DB 996 EPSKPT 1001
RESULT 9
CADD CHICK STANDARD; PRT; 771 AA.
AC P12957: 003698; 090756; 092061; 092018; 099230;
AD 01-JAN-1990 (Rel. 13, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Caldesmon (CDM).
GN CALD1 OR CAD.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RX TISSUE=9031;
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORM GIZZARD H-CAD).
RC TISSUE=Gizzard, and Oviduct;
RX MEDLINE=89340480; PubMed=2760048;
RA Bryan J., Imai M., Lee R., Moore P., Cook R.G., Lin W.-G.;
RT "Cloning and expression of a smooth muscle caldesmon.";
RN J. Biol. Chem. 264:13873-13879(1989).
[2]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORM GIZZARD H-CAD).
RC TISSUE=Gizzard;
RX MEDLINE=90026426; PubMed=2803315;
RA Hayashi K., Kanda K., Kimizuka F., Kato I., Sobue K.;
RT "Primary structure and functional expression of h-caldesmon
complementary DNA.";
RL Biochem. Biophys. Res. Commun. 164:503-511(1989).
[3]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORM BRAIN L-CAD).
RC TISSUE=Brain;
RX MEDLINE=91093148; PubMed=1824698;
RA Hayashi K., Fujio Y., Kato I., Sobue K.;
RT "Structural and functional relationships between h- and
l-caldesmons.";
RL J. Biol. Chem. 266:355-361(1991).
[4]
RP SEQUENCE FROM N.A. (GIZZARD H-CAD), BRAIN L-CAD AND GIZZARD L-CAD).
RX MEDLINE=94071934; PubMed=8250919;
RA Hatuna M., Hayashi K., Yano H., Takeuchi O., Sobue K.;
RT "Common structural and expressional properties of vertebrate
caldesmon genes.";
RL Biochem. Biophys. Res. Commun. 197:145-153(1993).
[5]
RP SEQUENCE FROM N.A. (ISOFORM GIZZARD L-CAD).
RC TISSUE=Gizzard;
RX MEDLINE=92042686; PubMed=1939602;
RA Bryan J., Lee R.;
RT "Sequence of an avian non-muscle caldesmon.";
RL J. Muscle Res. Cell Motil. 12:372-375(1991).
[6]
RP SEQUENCE OF 1-15 FROM N.A. (ISOFORM BRAIN L-CAD).
RC TISSUE=Brain;
RX MEDLINE=94271210; PubMed=8002994;
RA Yano H., Hayashi K., Hatuna M., Sobue K.;
RT "Identification of two distinct promoters in the chicken caldesmon
gene.";
RL Biochem. Biophys. Res. Commun. 201:618-626(1994).
[7]
RP SEQUENCE OF 451-756 FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Gizzard;
RX MEDLINE=89273666; PubMed=2730665;

RA Hayashi K., Yamada S., Kanda K., Kimizuka F., Kato I., Sobue K.;
RT "35 kDa fragment of h-caldesmon conserves two consensus sequences of
the tropomyosin-binding domain in troponin T.";
RL Biochem. Biophys. Res. Commun. 161:38-45(1989).
[8]
RP SEQUENCE OF 498-525.
RX MEDLINE=88293484; PubMed=3401222;
RA Morner D., Audemard E., Derancourt J.;
RT "Identification of a 15 kilodalton actin binding region on gizzard
caldesmon probed by chemical cross-linking.";
RL Biochem. Biophys. Res. Commun. 154:564-571(1988).
[9]
RP PHOSPHORYLATION SITES.
RX MEDLINE=92041815; PubMed=1939059;
RA Mak A.S., Carpenter M., Smillie L.B., Wang J.H.;
RT "Phosphorylation of caldesmon by p34cdc2 kinase. Identification of
phosphorylation sites.";
RL J. Biol. Chem. 266:19971-19975(1991).
[10]
RP PHOSPHORYLATION OF TYR-27 AND TYR-165.
RX MEDLINE=20026923; PubMed=10559276;
RA Wang Z., Danielson A.U., Mainle N.J., McManus M.J.;
RT "Tyrosine phosphorylation of caldesmon is required for binding to the
Shc.Grb2 complex.";
RL J. Biol. Chem. 274:33807-33813(1999).
[11]
RP FUNCTION: ACTIN- AND MYOSIN-BINDING PROTEIN IMPLICATED IN THE
REGULATION OF ACTOMYOSIN INTERACTIONS IN SMOOTH MUSCLE AND
NONMUSCLE CELLS (COULD ACT AS A BRIDGE BETWEEN MYOSIN AND ACTIN
FILAMENTS). STIMULATES ACTIN BINDING OF TROPOMYOSIN WHICH
INCREASES THE STABILIZATION OF ACTIN FILAMENT STRUCTURE. IN MUSCLE
TISSUES, INHIBITS THE ACTOMYOSIN ATPASE BY BINDING TO F-ACTIN.
THIS INHIBITION IS ATTENUATED BY CALCIUM-CALMODULIN AND IS
POTENTIATED BY TROPOMYOSIN. INTERACTS WITH ACTIN, MYOSIN, TWO
MOLECULES OF TROPOMYOSIN AND WITH CALMODULIN. ALSO PLAY AN
ESSENTIAL ROLE DURING CELLULAR MITOSIS AND RECEPTOR CAPING.
[12]
RP SUBCELLULAR LOCATION: ON THIN FILAMENTS IN SMOOTH MUSCLE AND ON
STRESS FIBERS IN FIBROBLASTS (NONMUSCLE) (BY SIMILARITY).
[13]
RP ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=3;
Name=Gizzard h-cad;
Isoid=P12957-2; Sequence=Displayed;
Name=Brain l-cad;
Isoid=P12957-1; Sequence=Displayed;
Name=Gizzard l-cad;
Isoid=P12957-3; Sequence=VSP_004153;
[14]
RP TISSUE SPECIFICITY: HIGH-MOLECULAR-WEIGHT CALDESMON (H-CALDESMON)
IS PREDOMINANTLY EXPRESSED IN SMOOTH MUSCLES, WHEREAS LOW-
MOLECULAR-WEIGHT CALDESMON (L-CALDESMON) IS WIDELY DISTRIBUTED IN
NON-MUSCLE TISSUES AND CELLS. NOT EXPRESSED IN SKELETAL MUSCLE OR
HEART.
[15]
RP DOMAIN: THE N-TERMINAL PART SEEMS TO BE A MYOSIN/CALMODULIN-
BINDING DOMAIN, AND THE C-TERMINAL A TROPOMYOSIN/ACTIN/CALMODULIN-
BINDING DOMAIN. THESE TWO DOMAINS ARE SEPARATED BY A CENTRAL
HELICAL REGION IN THE MUSCLE FORMS.
[16]
RP PTM: Phosphorylated in non-muscle cells. Phosphorylation by CDC2
during mitosis causes caldesmon to dissociate from microfilaments.
Phosphorylation reduces caldesmon binding to Actin, Myosin, and
Calmodulin as well as its inhibition of actomyosin ATPase
activity. Phosphorylation also occurs in both quiescent and
dividing smooth muscle cells with similar effects on the
interaction with Actin and Calmodulin and on microfilaments
reorganization (by similarity).
[17]
RP SIMILARITY: BELONGS TO THE CALDESOM FAMILY.
[18]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
[19]
RP EMBL: J04968; AAA49067.1; -

OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION. -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: X17102; CAA34958.1; -.
 DR EMBL: X05558; CAA29073.1; -.
 DR PIR: S15762; S15762.
 DR InterPro: IPR006821; Filament_head.
 DR InterPro: IPR001664; IF.
 DR InterPro: IPR002957; Keratin_1.
 DR Pfam: PF000038; filament; 1.
 DR Pfam: PF04723; filament head; 1.
 DR PRINTS: PR01248; TYPEKERATIN.
 DR PROSITE, PS00226; IF; 1.
 KM Intermediate filament; Coiled coil; Neurone; Phosphorylation;
 Glycoprotein.
 FT INIT MET 0 0 HEAD.
 FT DOMAIN 1 98 ROD.
 FT DOMAIN 99 406 ROD.
 FT DOMAIN 407 857 TAIL.
 FT DOMAIN 99 130 COIL 1A.
 FT DOMAIN 131 143 LINKER 1.
 FT DOMAIN 144 242 COIL 1B.
 FT DOMAIN 243 259 LINKER 12.
 FT DOMAIN 260 281 COIL 2A.
 FT DOMAIN 282 285 LINKER 2.
 FT DOMAIN 286 406 COIL 2B.
 FT CARBOHYD 46 46 O-LINKED (GLCNAC) (BY SIMILARITY).
 FT CARBOHYD 426 426 O-LINKED (GLCNAC) (BY SIMILARITY).
 FT CONFLICT 546 546 G -> R (IN REF. 2).
 SQ SEQUENCE 857 AA; 95704 MW; 4E2B0FC6C64778B CRC64;

Query Match 11.5%; Score 292.5; DB 1; Length 857;
 Best Local Similarity 25.9%; Pred. No. 0.00017;
 Matches 130; Conservative 66; Mismatches 19; Indels 107; Gaps 18;

QY 35 HATENEGATVPPISSNPNANESQAEQGE-OPKLDSSRDARKKEVEEYKKIYGEYSYAKS 92
 Db 290 YAKLTAEAEONKEAIRSAKEEIAEYRROLQSKSIELESVGTGKSLERQSDIEERHND 149
 QY 93 TKRHITTVLVNVLNNIKRYEYLNKIYESTSESOLQIMESRKKVDNAVSKPEK----- 147
 Db 350 LTTYQDTIHOLENELRGTQKEMARHLEYPD-----LNVKXALDIEIAVYRKLLEGE 402
 QY 148 DSSSSSSDSTKPEASDTAKPNKPTPEKEVAEA---KKKVEBAKAKAQKEDRRNYP 204
 Db 403 ETRFSAASGSGITGPIFH-----RQPSVTIASTKIQTKIIEPKK----- 442
 QY 205 TITYKTELEIASDVEVKAEL-----ELVKVANEPRDEQIKQAE---EYESK 253
 Db 443 KVOHKFVEEIIETKVEDEKSEMEDALSAIAEEMAKAQEEOEKEEAEVEEAVSE 502
 QY 254 QAEATRIKKITKTREEEAEAEKARADAKBOGKPKGRKRGVPGELATPPDKENDAKSDS 313
 Db 503 KAAEQAAEEBEKEEAE 545
 QY 314 SVGEETLPSPSLKPEKVAEAEKVEBAKKAEDQKEDRRNYPNTYKTELEIASDV 373
 Db 546 --GEEA-----EEBEAEAKGAEEAGAKVEKVSPPAKSPKSPKSPKSPKSPKSPKSPK 592
 QY 374 EVKKAELVYEAEKPRNEEKVQAQAEVSKKA--EATRLKIKITDRKKAEEBAKKA 431
 Db 593 AVQKAAAEVGDQDAEKAEKAAKEEKAAKEEKAAKEEKAAKEEKAAKEEKAAKEEKAAKEEK 649

QY 432 AEEDKVK--EKPAEQOPAPAPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 489
 Db 650 ITPKVAASPEKPT-TEPKVSP--EKPAEPEKPTPEKPAEPEKPA----- 692

QY 490 EYNRLQOOPKTEKPAOPSTP 511
 Db 693 -----TEPKRTEKPAKPAKPEK 709

RESULT 11
 TRDN CANFA
 ID TRDN CANFA STANDARD; PRT; 700 AA.
 AC P82179;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Triadin.
 GN TRDN.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
 RC TISSUE=Heart, and Skeletal muscle;
 RX MEDLINE=99428545; PubMed=10497235;
 RA Kobayashi Y.M., Jones L.R.;
 RT "Identification of triadin 1 as the predominant triadin isoform expressed in mammalian myocardium."
 RL J. Biol. Chem. 274:28660-28668(1999).
 CC -1- FUNCTION: MAY BE INVOLVED IN ANCHORING CALSEQUESTRIN TO THE JUNCTIONAL SARCOPASMIC RETICULUM AND ALLOWING ITS FUNCTIONAL COUPLING WITH THE RYANODINE RECEPTOR (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Sarcoplasmic reticulum.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Name=Skeletal;
 CC IsoId=P82179-1; Sequence=Displayed;
 CC Name=Cardiac 1;
 CC IsoId=P82179-2; Sequence=VSP_004001, VSP_004002;
 CC Name=Cardiac 3;
 CC IsoId=P82179-3; Sequence=VSP_004003, VSP_004004;
 CC -1- TISSUE SPECIFICITY: SKELETAL AND CARDIAC MUSCLE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: AF165916; AAF00222.1; -.
 DR EMBL: AF165915; AAF00221.1; -.
 DR EMBL: AF165917; AAF00223.1; -.
 KW Transmembrane; Sarcoplasmic reticulum; Glycoprotein;
 KM Alternative splicing.
 FT INIT MET 0 0
 FT DOMAIN 1 46
 FT TRANSMEM 47 67
 FT DOMAIN 68 700
 FT CARBOHYD 68 700
 FT CARBOHYD 616 716
 FT VARSPDIC 257 277
 FT
 FT
 FT VARSPDIC 278 700
 FT
 FT
 FT VARSPDIC 466 466
 FT
 FT
 FT VARSPDIC 524 579

BY SIMILARITY.
 CYTOPLASMIC (POTENTIAL).
 POTENTIAL.
 LUMENAL (POTENTIAL).
 N-LINKED (GLCNAC...) (POTENTIAL).
 N-LINKED (GLCNAC...) (POTENTIAL).
 DOYAFRWIMDMVHGLDLPFG -> GKHSSEVAGSGKRTLG
 KKQIQ (in isoform Cardiac 1).
 /FTId=VSP_004001.
 Missing (in isoform Cardiac 1).
 /FTId=VSP_004002.
 E -> EPIKQEVKVGSLKEKE (in isoform Cardiac 3).
 /FTId=VSP_004003.
 BEKVVKQVATTEAAIEKTVKPKPAKKAHQEKESPTIKTD

```

FT FT KPKTSKETPEVTES -> GILOVVAVLNCFLVQFOODE
FT ELNVESKVRMHVLSHPTSRSTSPILVSTTCRT (in
FT Isoform Cardiac 3).
FT /FTID=VSP_004004.
SQ SEQUENCE 700 AA; 78152 MW; F033E3AA1BE0C56 CRC64;

Query Match 11.5%; Score 291.5; DB 1; Length 700;
Best Local Similarity 25.2%; Pred. No. 0.00016;
Matches 138; Conservative 86; Mismatches 201; Indels 123; Gaps 26;

OY 37 TENEGATVPPTSSNANESQAEOGQP-----KTLDSERDKAREVEYKIVGESYA 90
DB 128 TDKGIEEPPLKQKEIHKEKAEKPERKJLAKVAHEKEKV--KEKSEKKATHKEKI 186
OY 91 KSTKRHTITVA-----LVNELNITKNEVLKIVSTSSQ--LQILMMESRKV 138
DB 187 EKKEPEPTKVAKEERKAKTEBKIKEVKGQKQEVKPTAAKVEVQTPPKAKEKEKE 246
OY 139 DEAVSKFE-KDS-----SSSSSSDSTKPEASDPAKPNKPTPEGEKVAEAKK 184
DB 247 TAAVAKHGOQKQYAFRCRYMIDMFVHGDLPPOSPLPPLPTVQASRPT--PASPTLEGE 305
OY 185 KVEEAKKAKDQKEDRRNYPTTYKTLLELAESDVEYKKAELLVKYKANEPRDQKIX 244
DB 306 --EEKKKAKEKK-----VTSETKKKE--KEDVKKKSDKQTAIDVEKKEPKAPETK 352
OY 245 QAEAEVESKQAEATRLTKIKITDRAEAEERAKRAAKQKPKGAKGVGELATPDKK 304
DB 353 QGTIKVVAQAA-----AKDKEDKDSKTKTPVEE--HKGKKQ-----BKK 392
OY 305 ENDAKSSSSVGEETLPSPSLKPEKKV--AEAEKVEE-----AKKAEQDQKEDRRNYPTN 359
DB 393 EKYVPAASSKKEHAPPS-----EKQVAKKTRAEFTSAASTKQAVGKKEEK----- 441
OY 360 TYKTLLELEI-----AESDVEYKKAELLVKEBAKPRNEEKYKQAKAEVESKKAETRL 414
DB 442 TTKTVEQEIIRKEKSGKSTASQKPEIKIDE--KMPKADKEVKPKRPOSQVKEKSE--S 499
OY 415 KIKDKRK-----KAE-----EAKRAAEDEKVEKEAE--OPQAP 449
DB 500 QVKEAKEBQDIAPKPTVSHGKPEKVKVQVAKATEKAIETVAPKPAKKAHEHOKESP 559
OY 450 APKAEKPAAPK--PE-----NPAPQKAEKPADQQAEDVYARSEEEYNR--LTQOQPPK 501
DB 560 TIKDKPKPKTSKETPEVTESGKKKIEKSEKSEKKAEMKHLKEEKVSTRKESLSQSHNTX 619
OY 502 TEKPAOPS 509
DB 620 AEKPARVS 627

RESULT 12
TRDN HUMAN STANDARD; PRT; 728 AA.
ID TRDN HUMAN Q13061;
AC 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Triadin.
GN TRDN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=96061957; PubMed=7588753;
RA Taake N.L., Eyre H.T., O'Brien R.O., Sutherland G.R., Denborough M.A.,
RA Foster P.S.;
RT "Molecular cloning of the cDNA encoding human skeletal muscle triadin
RT and its localisation to chromosome 6q22-6q23.";
RL Eur. J. Biochem. 233:258-265 (1995).

```

```

CC CC -!- FUNCTION: MAY BE INVOLVED IN ANCHORING CALDESQUESTRIN TO THE
CC JUNCTIONAL SARCOPLASMIC RETICULUM AND ALLOWING ITS FUNCTIONAL
CC COUPLING WITH THE RYANODINE RECEPTOR (BY SIMILARITY).
CC -!- SUBUNIT: Homooligomer of variable subunit number; disulfide-linked
CC (by similarity).
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Sarcoplasmic
CC reticulum.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U18985; AAA75315.1; -.
CC PIR; S68191; S68191.
CC DR GeneW; HGNC:12261; TRDN.
CC MIM; 603283; -.
CC DR GO; GO:0016021; C:integral to membrane; TAS.
CC DR GO; GO:0006936; P:muscle contraction; TAS.
CC KW Transmembrane; Sarcoplasmic reticulum; Glycoprotein.
CC FT INIT MET 0
CC FT DOMAIN 1 46
CC FT TRANSMEM 47 67
CC FT POTENTIAL.
CC FT DISULFID 269 269
CC FT DISULFID 690 690
CC FT CARBOHYD 74 74
CC FT CARBOHYD 646 646
CC FT N-LINKED (GLCNAC...) (POTENTIAL).
CC FT SEQUENCE 728 AA; 81423 MW; C1C53BBE1B2A0815 CRC64;

Query Match 11.4%; Score 290; DB 1; Length 728;
Best Local Similarity 23.0%; Pred. No. 0.00016;
Matches 126; Conservative 95; Mismatches 208; Indels 118; Gaps 21;

OY 38 ENEGATVPPTSSNANESQAEOGQPKKLDSEBDKAREVEYKIVGESYASTKKRH 97
DB 154 ERKIQTVTHKEKKGKREKREKPKAKTHKEIKKEKEPEPTKY-----AEQKKAK 208
OY 98 TITVALVNELNITKNEVLKIVESTS--SOLQILMMESRSKVD--AVSKFEKSSSSS 153
DB 209 TAEKSEKTKKEVKGQKQEVKQTAARKEVQTPSKPEKEDKKAASVGHQKQYAF 268
OY 154 SS-----DSSTPEASDPAKPNKPTPEGEKVAEAKKYVEAKKADQKEDRRNYPTI 206
DB 269 CRYMIDIFVHGDLPQSPALPPLPTQASRPTPASPALEE--KEGKKAKEK-----V 322
OY 207 TYKTLLELAESDVEYKKAELLVKYKANEPRDQKIKQAEAEVESKQAEATRLKIKITD 266
DB 323 TSETKKKE--KEDIKKSSSEKTAIDVEKKEPKGASSETKQIVTKIAQAA-----AK 371
OY 267 REAEAEAKRRADAEQKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPSPSLK 326
DB 372 KDEKKDSKTKKRAVEAQPGKKQ-----EKKEKVEBAKSPKKEHSVPS----- 417
OY 327 PEKV--AEAEKVEE-----AKKAEQDQKEDRRNYPTNTYKTLLELAE-----S 371
DB 418 -DKQVKAATERAKEIEAVSSKKAVPGKKEK-----TTKTVQEIIRKEKSGTSSILK 470
OY 372 DVENKKALELV-----KEBAKPRNEEKVQAKAEVESK-----KAEATRLKIKITDK 421
DB 471 DKPEPIKGEKVPASLLEKEPEYTKDKKMSKAGKVEKVRPPQLOGKKEKPEPKKEAK 530
OY 422 KAESE-----AKKAAEDBKVEKPAEQOPAPAPAEKPAAPKPNPAEQP-- 469
DB 531 PAISEKQIHKQDILVREKTVSHGKPREKYLKQKAVATIEKTAQPKTKKAEHNERBPS 590
OY 470 -KAEP-----ADQQAEE--DYARSEEEYN-----RLTQOQPPKT 502
DB 591 IKTKPKRPTPKTSKETPEVTESGKKKTEISEKSEKAKDKHILREKVSSTRKESLQAHNTYKA 650

```

QY	503	EXPAPQS	509
Db	651	EXPAPQS	657
RESULT	13		
ID	NFM_RABIT	STANDARD;	PRT; 644 AA.
AC	P54938;		
DT	01-OCT-1996 (Rel. 34, Created)		
DT	01-OCT-1996 (Rel. 34, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Neurofilament triplet M protein (160 kDa neurofilament protein)		
DE	(Neurofilament medium polypeptide) (NF-M) (fragment).		
GN	NEF3 OR NEFM OR NFM.		
OC	Oryctolagus cuniculus (Rabbit).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.		
OX	NCBI_Taxid=9986;		
RL	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Heart;		
RA	MEDLINE=97055255; PubMed=8899542;		
RT	Vitadello M., Vettore S., Lamar E., Chien K.R., Gorza L.;		
RT	"Neurofilament M mRNA is expressed in conduction system myocytes of		
RT	the developing and adult rabbit heart."		
RL	J. Mol. Cell. Cardiol. 28:1833-1844(1996).		
CC	-1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,		
CC	AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.		
CC	-1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFM IS		
CC	PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS		
CC	THOUGHT THAT PHOSPHORYLATION OF NFM RESULTS IN THE FORMATION OF		
CC	INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE		
CC	OF AXONAL CALIBER.		
CC	-1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING		
CC	OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE		
CC	LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND		
CC	COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.		
CC	-1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on		
CC	use. By non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; Z47378; CA87454.1; -		
DR	PIR; S55395; S55395.		
DR	InterPro; IPR001664; IF.		
DR	Pfam; PF00038; filament; 1.		
DR	PROSITE; PS00226; IF; 1.		
KW	Intermediate filament; Coiled coil; Neutone; Phosphorylation;		
KW	Glycoprotein.		
FT	NON_TER 1		
FT	DOMAIN 1		
FT	DOMAIN 197		
FT	DOMAIN 643		
FT	DOMAIN 198		
FT	DOMAIN <1		
FT	DOMAIN 33		
FT	DOMAIN 34		
FT	DOMAIN 50		
FT	DOMAIN 51		
FT	DOMAIN 72		
FT	DOMAIN 73		
FT	DOMAIN 76		
FT	DOMAIN 197		
FT	CARBOHYD 217		
FT	SEQUENCE 644 AA; 72450 MW; 030FDA642869676 CRC64;		
QY	57	AEQGEQPKLDSNDKARKEVEYVKRVGS-----YAKSKRKRTITV-----	101
Db	88	AEQNKQAIR-----SAKEELAEFRRLQSSISIEESVAMKESLERRVSDIEERHHNDL	141
Query Match	11.2%;	Score 284;	DB 1; Length 644;
Best Local Similarity	25.2%;	Pred. No. 0.00027;	
Matches 124;	Conservative 79;	Mismatches 167;	Indels 122; Gaps 20;

QY	102	-----ALVLELNINIKNEYLNKTVSESTSESQLOILIMMSSKSVDEAVSKEK-----D	148
Db	142	SSYQDTIQQLLEHLELGTCKEMARHLREYD-----LLNVGMALDIETLAATYKLLGEE	194
QY	149	SSSSSSSSSTKPEASDTAKPNKPTPEGKVAEAKKVE--EAKKAKDQKEEDRRNYPTIT	207
Db	195	TRFSFGSGITGLOPLTYH-----RQSVTISKSIOKTKVEAPKLT-----VQ	235
QY	208	YKTLLELEIAESVVEYKKALE-----LVYKANPDRBOQIKQAEVYESQALAE	257
Db	236	HKFVEEIIIEETVDEBKSMEBALTAIAEELAVSYEEKEEBAEKGEEQEAEEBVAA	295
QY	258	TRKKIKITDREAEAEAKRADAKQCKPKRAGKGVPELATTPODKENDAKSSDSVGE	317
Db	296	KKSPYKATTPPELKEEGEKEEGEQEEBEE--EDGYSVDQAEBOGSEKEGSSKN--GE	351
QY	318	ETLPEPSLKPEKKVAEAKKVAEAKKAKDQKEEDRRNYPTNYKTLLELEIAESDVYKK	377
Db	352	Q-----BEGTEAGEVEVEAEAKKEKTKTEKSE-----EVAKKEPPTYE	390
QY	378	AELLEVKEAKEKPRNEEYKQAKAEVSEKKAATYLE-KITDRKKAEEBAKRAAE---	433
Db	391	AKVG-KPEKASPVKSPVVEYK-----PKAATLKGEGQKEEVEYEEKKAAVESPK	444
QY	434	EDVYKEKAEQOPAPAPAKAEKPAAPKPENEA-----EQPAEKPADQOAE	481
Db	445	EKKV-EKKEEKXQDVPKKKAESPVVEEAAAEBAATTTKPTKVLKETTKEGKPLQOEKXK	503
QY	482	DYARR--SEEE	490
Db	504	EKAGEGSGSEEE	515

RESULT 14			
NFM_HUMAN	STANDARD;	PRT;	915 AA.
AC	P07197;		
DT	01-APR-1988 (Rel. 07, Created)		
DT	01-AUG-1988 (Rel. 08, last sequence update)		
DT	16-OCT-2001 (Rel. 40, last annotation update)		
DE	Neurofilament triplet M protein (160 kDa neurofilament protein)		
DE	(Neurofilament medium polypeptide) (NF-M) (Neurofilament 3).		
GN	NEF3 OR NEFM OR NFM.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=87275853; PubMed=3608989;		
RA	Myers M.W., Lazzarini R.A., Lee V.M.-Y., Schlaepfer W.W., Nelson D.L.;		
RT	"The human mid-size neurofilament subunit: a repeated protein		
RT	sequence and the relationship of its gene to the intermediate		
RT	filament gene family.";		
RL	EMBO J. 6:1617-1626(1987).		
RN	[2]		
RP	SEQUENCE OF ONE OF THE 13 RESIDUE REPEATS.		
RX	MEDLINE=88158120; PubMed=2450354;		
RA	Lee V.M.-Y., Otvos U. Jr., Carden M.J., Hollosi M., Dietzschold B.,		
RA	Lazzarini R.A.;		
RT	"Identification of the major multiphosphorylation site in mammalian		
RT	neurofilaments.";		
RT	Proc. Natl. Acad. Sci. U.S.A. 85:1998-2002(1988).		
RL	-I- FUNCTION. NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,		
CC	AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.		
CC	-I- PWM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P. NFM IS		
CC	PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS		
CC	THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF		
CC	INTERMEDIATE CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE		
CC	OF AXONAL CALIBER.		
CC	-I- PWM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING		
CC	OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE		

LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
 COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
 -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL; Y00067; CAA68276.1; -
 DR PIR; A27864; A27864.
 DR GeneW; HGNC:7734; NEF3.
 DR MIM; 162250; -
 DR GO; GO:0005883; C:neurofilament; TAS.
 DR GO; GO:0005200; F:structural constituent of cytoskeleton; TAS.
 DR InterPro; IPR006821; Filament_head.
 DR InterPro; IPR001664; IF.
 DR InterPro; IPR002957; Keratin_1.
 DR Pfam; PF00038; Filament; 1.
 DR Pfam; PF04732; Filament_head; 1.
 DR PRINTS; PR01248; TYPE1KERATIN.
 DR PROSITE; PS00226; IF; 1.
 KW Intermediate filament; Coiled coil; Neutrone; Phosphorylation;
 KW Glycoprotein.
 FT INIT_MET 0
 FT DOMAIN 1 103 HEAD.
 FT DOMAIN 104 411 ROD.
 FT DOMAIN 412 915 TAIL.
 FT DOMAIN 104 135 COIL_1A.
 FT DOMAIN 136 148 COIL_1B.
 FT DOMAIN 149 247 COIL_1B.
 FT DOMAIN 248 264 LINKER_12.
 FT DOMAIN 265 286 COIL_2A.
 FT DOMAIN 287 290 LINKER_2.
 FT DOMAIN 291 411 COIL_2B.
 FT DOMAIN 613 690 6 X 13 AA TANDEN REPEATS.
 FT CAROHD 46 46 O-LINKED (GLCNAC) (BY SIMILARITY).
 FT CAROHD 430 430 O-LINKED (GLCNAC) (BY SIMILARITY).
 SQ SEQUENCE 915 AA; 102317 MW; 5F5D3DF34CD9E50 CRC64;
 Query Match 11.2%; Score 283; DB 1; Length 915;
 Best Local Similarity 24.0%; Pred. No. 0.0004;
 Matches 118; Conservative 85; Mismatches 199; Indels 90; Gaps 16;
 QY NESQAEGEOPKLDSEEDKARK-----EVEEYVKTIAGESYAKST-----KKR 96
 DB 367 NEIQRKTQKMAHREYQDLNVKMAIDIEIAAYRKLLEGERTSTFAGSTGPLYTHR 426
 QY HTITVALVNEIN-----NIKETLAKIVESTSSQILMERSKVDENVKFEEDS 149
 DB 427 PPIITSSIKIOKTVKAPKLVQHKFVEEIIETKVEDKSEMEELAITTELAASKEE 486
 QY SSSSSSDSTPEAD--TAK--PNKPTPEKVAEKKVVEAKKADQKEEDRRYP 204
 DB 487 KKEAEKEEPEAEVEEVAAKSVKATAPVKEEKEEKEEKEEKEEKEEKEEKEE 544
 QY TITVTELELEIAESDVEVKALELVKVKANFPDEOKIKQAEVESKQAEATLEKIK 264
 DB 545 -----DQAEEGSEKEGSE-----KEBEGEGEGTEAEABGEAEKKEKVE 588
 QY 265 TDREBAEAEAKRADAKQKPKGAKGVPEGLATPDKENDAKSSSVGEETLPSPS 324
 DB 589 EKSEVATKEEIVADAKVE--KPE--KAKSPVPKSPVPEEKGKSPVPKSP 646
 QY 325 LKPEKVAEAKKVE--AKKKAEDQKEDRRNYPNTYKTELEIAESDVEVKALELV 383
 DB 647 VEEKGKSPVPSVPEEKGKSPVSKAKSP-----VPKSPVEAKSKAEVG 696
 QY 384 KEEAKEPNEEKVKQAKAEVSKAEATRLKIK--TDKKAEEBAKAKAA----- 432

DB 697 KGEQKE--EEKEVKEAPKEEVEKKE-----EKPDVEKKAESPVKEAEVAEVYITKS 751
 QY 433 -----EEDDKKEKPAEDPOPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 486
 DB 752 VKVHLEKTKKEEGKPLQO-----EKKEKAGGEGSEEGSDKAGKSKREDIAVN 802
 QY 487 SEEEYNRLTQOQ 498
 DB 803 GEVEKKEVEOS 814
 RESULT 15
 NFM_MOUSE STANDARD; PRT; 848 AA.
 AC P08553; Q61961; -
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Neurofilament triplet M protein (160 kDa neurofilament protein)
 DE (Neurofilament medium polypeptide) (NF-M).
 GN NEF3 OR NEFM OR NFM.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=8724694; PubMed=3036526;
 RA Levy E., Liem R.K.H., D'Eustachio P., Cowan N.J.;
 RT "Structure and evolutionary origin of the gene encoding mouse NF-M,
 RL the middle-molecular-mass neurofilament protein";
 Eur. J. Biochem. 166:71-77(1987).
 [2]
 RP SEQUENCE OF 322-540 FROM N.A.
 RX MEDLINE=8715637; PubMed=3103856;
 RA Julien J.-P., Meyer D., Flavell D., Hurst J., Grosved F.;
 RT "Cloning and developmental expression of the murine neurofilament
 RL gene family";
 Brain Res. 387:243-250(1986).
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 CC -1- THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFM IS
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
 CC THOUGHT THAT PHOSPHORYLATION OF NFM RESULTS IN THE FORMATION OF
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
 CC OF AXONAL CALIBER.
 CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL; X05640; CAA29127.1; -
 DR EMBL; M20481; AAA39815.1; -
 DR PIR; B43772; B43772.
 DR PIR; B43772; B43772.
 DR MGI; MGI:97314; NeF3.
 DR InterPro; IPR006821; Filament_head.
 DR InterPro; IPR001664; IF.
 DR Pfam; PF00038; Filament; 1.
 DR Pfam; PF04732; Filament_head; 1.
 DR PROSITE; PS00226; IF; 1.
 KW Intermediate filament; Coiled coil; Neutrone; Phosphorylation;
 KW Glycoprotein.
 FT INIT_MET 0 0

```
FT DOMAIN 1 102 HEAD.
FT DOMAIN 103 410 ROD.
FT DOMAIN 411 848 TAIL.
FT DOMAIN 103 134 COIL 1A.
FT DOMAIN 135 147 LINKER 1.
FT DOMAIN 148 246 COIL 1B.
FT DOMAIN 247 263 LINKER 12.
FT DOMAIN 264 285 COIL 2A.
FT DOMAIN 286 289 LINKER 2.
FT DOMAIN 290 410 COIL 2B.
FT CARBOHYD 47 47 O-LINKED (GLCNAC) (BY SIMILARITY).
FT CARBOHYD 430 430 O-LINKED (GLCNAC) (BY SIMILARITY).
FT CONFLICT 432 432 S -> F (IN REF. 2).
FT CONFLICT 539 540 QA -> RR (IN REF. 2).
SQ SEQUENCE 848 AA; 95910 MM; 5F251F274D0F13B6 CRC64;
```

Query Match 11.1%; Score 282.5; DB 1; Length 848;
Best Local Similarity 24.9%; Pred. No. 0.00039;
Matches 133; Conservative 79; Mismatches 216; Indels 107; Gaps 21;

```
QY 35 HATNEGATQVPTSNRANESQAGE--QPKLDSERDKARKEVEEYVKI VGSYAKS 92
DB 294 YAKLTEAEQNKAEIRSAKEIEAYRQLQSKSIELESVGTESLERQLSDIERHND 353
QY 93 TKRHTITVALVNEIINKNEYLINKIVESTESQQLIMMESRSKVDKAVSKFEK----- 147
DB 354 LSSYODTIQOLENELRGTKWEMARLREYQD-----LINVKALDIETIAVRKLEGE 406
QY 148 DSSSSSSSSDST-----KPEASPTAKPNKPTPEGEVAAKKEVEAKKADQKEEDR 200
DB 407 ETRFSTFGSITGPLYTHRQPSVTISISKIQTKEAPKLKVQHKFVEI--IEETKVEDE 464
QY 201 RNYPTITKTELELEIAESDVEVKALELVKANKNEPRDEQK-IKQAEAVESKQAEATR 259
DB 465 KSEMEETITATAIEBLAAS---AKEEKEAEKEEPEAKSPVKSPEAKEEESGE--- 517
QY 260 LKKITDREAEAEAEAKR-----RADAKEQKPKGRAGVPGELATPDKENDAKSSDS 314
DB 518 ---KEEBEGQHEEBEGVKSQAEQ---GSEKEG--SSEKDEGQHEEBEGTEAEG 569
QY 315 VGEETLPSPSLKPEKVAEAKKVEAKKADQKEEDRNYPTNTYKTELEIAESDVE 374
DB 570 EGEEA---EAKKEEKI---EGKVEEVAVKEEIKVEKP-----EKAKSPMP 608
QY 375 VKKAELEIVKEEK---EPNNEKVQAKAVESKKAETRLKIK-----TDKKAE 424
DB 609 KSPVEEVKPKPEAKAGKEQKEEKEVEEKEKVTKEKPEKEKVEKEKPPROVADKKAE 668
QY 425 EEAERKAAEF-----DKVK---EKPAEQOPAPAPAKAPAPKPPENPAEQPAEKPAD 476
DB 669 SPVKEKAVEEYITTSKSVKLEKDTKEKQPOKEKVEKA---EEBGSSEBGSDRSPQ 725
QY 477 QQAEDYARRSE---EEYNRLTQQQ-----PPKTEKPAQPS 510
DB 726 ESKKEDIANGVEVGEKEEEOETQEKSGREBEKGVVTNGLDVSPAEEKGEDSS 780
```

Search completed: November 21, 2003, 13:32:32
Job time : 17.7902 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:22:21 ; Search time 55.4631 Seconds
(without alignments)
1462.401 Million cell updates/sec

Title: US-09-298-523C-13
Perfect score: 2537
Sequence: 1 MFAKSRKHYHYSIRKFSVG.....NRLTQQQPKTKPKAQSTP 511

Scoring table: BIOSUM62
Gap0 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

A Geneseq 19Jun03:*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2516	99.2	693	23	AAU76890
2	2516	99.2	693	24	ABU02720
3	2516	99.2	694	21	AAV81653
4	2350.5	92.6	655	20	AAV49226
5	2350.5	92.6	655	20	AAV12039
6	2252.5	88.8	460	20	AAV49250
7	2249.5	88.7	460	20	AAV32189
8	2235.5	88.1	459	20	AAV49251
9	2235.5	88.1	459	20	AAV32190

10	2194.5	86.5	446	20	AAV49143	Amino acid sequenc
11	2177.5	85.8	446	20	AAV49140	Amino acid sequenc
12	2101.5	82.8	428	20	AAV49238	Choline binding pr
13	2101.5	82.8	428	20	AAV32110	Choline binding pr
14	1979.5	78.0	406	20	AAV49225	Polypeptide R2 of
15	1979.5	78.0	406	20	AAV32098	Choline binding pr
16	1770.5	69.8	487	23	AAU76889	Pneumococcal surfa
17	1730	68.2	564	19	AAV62654	C3 binding protein
18	1726	68.0	701	23	AAU76888	Pneumococcal surfa
19	1718	67.7	581	20	AAV43394	S. pneumoniae SpSA
20	1609.5	63.4	488	23	AAU76887	Pneumococcal surfa
21	1569.5	61.9	663	20	AAV49232	ChpA of serotype R
22	1569.5	61.9	663	20	AAV32104	Choline binding pr
23	1569.5	61.9	663	20	AAV43395	S. pneumoniae chol
24	1569.5	61.9	663	23	AAE29201	Streptococcus pneu
25	1569.5	61.9	663	23	ABG60639	Partial sequence o
26	1569.5	61.9	663	24	ABP55319	Streptococcus pneu
27	1568.5	61.8	419	20	AAV49243	N-terminal region
28	1568.5	61.8	419	20	AAV32182	N-terminal region
29	1512.5	59.6	406	20	AAV49152	Amino acid sequenc
30	1506	59.4	631	18	AAV46444	ChpA, a choline bi
31	1504.5	59.3	439	20	AAV49240	N-terminal region
32	1504.5	59.3	439	20	AAV32179	N-terminal region
33	1497.5	59.0	437	20	AAV49244	N-terminal region
34	1497.5	59.0	437	20	AAV32183	N-terminal region
35	1496.5	59.0	437	20	AAV49241	N-terminal region
36	1496.5	59.0	437	20	AAV32180	N-terminal region
37	1494	58.9	439	20	AAV49242	N-terminal region
38	1494	58.9	439	20	AAV32181	N-terminal region
39	1466.5	57.8	439	20	AAV49249	N-terminal region
40	1466.5	57.8	439	20	AAV32188	N-terminal region
41	1465.5	57.8	433	20	AAV49245	N-terminal region
42	1465.5	57.8	433	20	AAV32184	N-terminal region
43	1455.5	57.4	426	20	AAV49146	Amino acid sequenc
44	1447.5	57.1	425	20	AAV49147	Amino acid sequenc
45	1446.5	57.0	424	20	AAV49148	Amino acid sequenc

ALIGNMENTS

RESULT 1	AAU76890	standard; Protein; 693 AA.
XX	AAU76890;	
AC	05-JUN-2002	(first entry)
XX		
DE	Pneumococcal surface protein C19 TIGR (PspC19 TIGR).	
XX	Hic; factor H; FH; antibacterial; PspC19 TIGR;	
KW	pneumococcal surface protein C19 TIGR.	
KM		
XX		
OS	Streptococcus pneumoniae.	
XX		
XX	W0200208426-A2.	
PN		
XX	31-JAN-2002.	
PD		
XX		
PF	19-JUL-2001; 2001WO-EP08409.	
XX		
XX	20-JUL-2000; 2000SE-0002738.	
FR		
XX	(HANS-) HANSA MEDICAL AB.	
PA		
XX	Bjorck L, Sjöholm A, Janulczyk R, Pozzi G, Iannelli F;	
PI	WPI, 2002-257337/30.	
XX		
XX	Polypeptide having ability to bind factor H, and proteins and peptides	
PT	derived from polypeptides useful as vaccine for infections caused by	
PT	bacteria and for identifying agents inhibiting binding of factor H to	

[illegible][illegible]


```

Db 121 ESTSESQIIMMSRSKVDKAVSKFEKDSSSSSDSTKEASDTAKPNKPTPEGKV 180
QY 180 AEAKKKVEEA-KKAKDKQEDRRNPTITTYKLELEIAESDVEVKKALELVKYNANPR 238
Db 181 AEAKKKVEEAERKAKDKQEDRRNPTITTYKLELEIAESDVEVKKALELVKYNANPR 240
QY 239 DEOKIKQAEAVESKQAEATRLKIKITDREAEAEAKRRADAKQEGKPKGAKGVPEGL 298
Db 241 DEOKIKQAEAVESKQAEATRLKIKITDREAEAEAKRRADAKQEGKPKGAKGVPEGL 300
QY 299 ATPDKKENDAKSSDSSVGEETLPSPSLKPEKKVAEAKKVEBAKKKADQKEEDRRNPT 358
Db 301 ATPDKKENDAKSSDSSVGEETLPSPSLKPEKKVAEAKKVEBAKKKADQKEEDRRNPT 360
QY 359 NTYKTLLELEIAESDVEVKKALELVKEAKEPRNEEKYQAKAVESKKAATRLKIKIT 418
Db 361 NTYKTLLELEIAESDVEVKKALELVKEAKEPRNEEKYQAKAVESKKAATRLKIKIT 420
QY 419 DRKKAEEBAKRAAEEDVKKEKPAEQOPAPAPAKAEKPAPAKPENPAEQPKAEKPADQ 478
Db 421 DRKKAEEBAKRAAEEDVKKEKPAEQOPAPAPAKAEKPAPAKPENPAEQPKAEKPADQ 480
QY 479 AEDYARRSEEEYNRLTQOQPPKTEKPAQPSPT 511
Db 481 AEDYARRSEEEYNRLTQOQPPKTEKPAQPSPT 513

```

RESULT 3
AAV81653
ID AAV81653 standard; Protein; 694 AA.

AC AAV81653;

DT 24-MAY-2000 (first entry)

XX Streptococcus pneumoniae protein sequence ID302.

XX Streptococcus pneumoniae; vaccine; screening; protein antigen;

KM antibacterial; antinflammatory; meningitis; infection; diagnosis;

XX Streptococcus pneumoniae.

OS Streptococcus pneumoniae.

XX WO200006737-A2.

XX 10-FEB-2000.

XX 27-JUL-1999; 99WO-GB02451.

XX 27-JUL-1998; 98GB-0016337.

XX 19-MAR-1999; 99US-0125164.

XX (MICR-) MICROBIAL TECHINCS LTD.

XX Gilbert CPG, Hansbro PM;

XX WPI; 2000-195300/17.

XX New Streptococcal protein, useful as a vaccine, for diagnosis of

PT pneumococcal diseases and for screening agents capable of antagonizing

PT or inhibiting expression of the protein

XX Claim 2; Page 96; 108pp; English.

CC AAV81501 to AAV81679 represent specifically claimed protein sequences

CC isolated from Streptococcus pneumoniae. AAV81501 to AAV81679 represent

CC specifically claimed nucleotide sequences isolated from S. pneumoniae.

CC The sequences have antibacterial and antinflammatory properties.

CC The protein sequences, and fragments of them, are useful as immunogens

CC and/or antigens. The nucleotide sequences can be used in vaccines and in

CC diagnostic assays. The proteins and nucleotides can be useful for the

CC detection and diagnosis of S. pneumoniae. The protein sequences are also

CC useful for screening an agent capable of antagonizing, inhibiting or

CC interfering with the function or expression of the proteins in which the

CC agent is useful for treatment or prophylaxis of S. pneumoniae infection

CC and meningitis. AAV81501 to AAV81679 represent primers used in the

CC exemplification of the present invention.

CC

XX

SQ Sequence 694 AA;

Query Match 99.2%; Score 2516; DB 21; Length 694;

Best Local Similarity 99.6%; Pred. No. 3,76-152;

Matches 511; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

1 MFASKSRKHYHSIRKFSVG-ASVVAASLVWGSVHAATENGCATQVPTSSRNANESQAEQ 59

1 MFASKSRKHYHSIRKFSVGASVVAASLVWGSVHAATENGCATQVPTSSRNANESQAEQ 60

60 GEOPKLDSEDRKARKEVEEYVKKI VGESYAKSTKKHTITVALVNEIINKNEYLNKIV 119

61 GEOPKLDSEDRKARKEVEEYVKKI VGESYAKSTKKHTITVALVNEIINKNEYLNKIV 120

120 ESTSESQIIMMSRSKVDKAVSKFEKDSSSSSDSTKEASDTAKPNKPTPEGKV 179

121 ESTSESQIIMMSRSKVDKAVSKFEKDSSSSSDSTKEASDTAKPNKPTPEGKV 180

180 AEAKKKVEEA-KKAKDKQEDRRNPTITTYKLELEIAESDVEVKKALELVKYNANPR 238

181 AEAKKKVEEAERKAKDKQEDRRNPTITTYKLELEIAESDVEVKKALELVKYNANPR 240

239 DEOKIKQAEAVESKQAEATRLKIKITDREAEAEAKRRADAKQEGKPKGAKGVPEGL 298

241 DEOKIKQAEAVESKQAEATRLKIKITDREAEAEAKRRADAKQEGKPKGAKGVPEGL 300

299 ATPDKKENDAKSSDSSVGEETLPSPSLKPEKKVAEAKKVEBAKKKADQKEEDRRNPT 358

301 ATPDKKENDAKSSDSSVGEETLPSPSLKPEKKVAEAKKVEBAKKKADQKEEDRRNPT 360

359 NTYKTLLELEIAESDVEVKKALELVKEAKEPRNEEKYQAKAVESKKAATRLKIKIT 418

361 NTYKTLLELEIAESDVEVKKALELVKEAKEPRNEEKYQAKAVESKKAATRLKIKIT 420

419 DRKKAEEBAKRAAEEDVKKEKPAEQOPAPAPAKAEKPAPAKPENPAEQPKAEKPADQ 478

421 DRKKAEEBAKRAAEEDVKKEKPAEQOPAPAPAKAEKPAPAKPENPAEQPKAEKPADQ 480

479 AEDYARRSEEEYNRLTQOQPPKTEKPAQPSPT 511

481 AEDYARRSEEEYNRLTQOQPPKTEKPAQPSPT 513

RESULT 4

AAV49226

ID AAV49226 standard; Protein; 655 AA.

XX AAV49226;

XX 07-FEB-2000 (first entry)

XX CBPA of serotype 4 amino acid sequence.

XX Choline binding protein A; CBPA; truncate; immune response; infection;

XX Streptococcus pneumoniae.

PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
XX
XX Tuomanen EI, Maure HR;
PI
XX WPI, 1999-620161/53.
DR N-PSDB; AA231402.
XX
XX Novel polypeptides, used to develop products for the diagnosis,
PT prevention and treatment of pneumococcal infections -
XX
XX Disclosure; Page 9; 85pp; English.
PS
XX
XX The invention provides novel isolated polypeptides comprising the amino
CC acid sequence of an N-terminal choline binding protein A (Cbpa) truncate
CC (CbpaT). The polypeptides can be selected from sequences shown in
CC AA19225, AA19227, AA19230, AA19231, AA19233 and AA19238. The
CC polypeptides can be used for inducing an immune response in a subject
CC which has been exposed to or infected with a pneumococcal bacterium. They
CC can also be used for preventing infection by a pneumococcal bacterium.
CC Vaccines comprising the polypeptides or encoding nucleic acids can be
CC used for treating a subject infected with or exposed to a pneumococcal
CC bacterium. Antibodies specifically binding the polypeptides can be used
CC for detection and diagnosis and for preventing pneumococcal attachment to
CC a mucosal surface. The products can be used in humans and other animals
CC such as domestic animals, such as feline or canine subjects, farm animals
CC such as bovine, equine, caprine, ovine, and porcine subjects, wild
CC animals (whether in the wild or in a zoological garden), research
CC animals, such as mice, rats, rabbits, goats, sheep, pigs, dogs, cats,
CC i.e. for veterinary medical use.
XX
XX Sequence 655 AA;
SQ

Query Match 92.6%; Score 2350.5; DB 20; Length 655;
Best Local Similarity 99.8%; Pred. No. 1.2e-141;

Matches 474; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 38 ENEGATQVPTSSNRANESQAQEQOPKLDSEBDKARKEVEEYVKIIGESYAKSTKKRH 97
DB 1 ENEGATQVPTSSNRANESQAQEQOPKLDSEBDKARKEVEEYVKIIGESYAKSTKKRH 60
QY 98 TITVALVNLNINKNEYLINKIVESTESQLOILMESRSKYDAVSKFEKSSSSSSSDS 157
DB 61 TITVALVNLNINKNEYLINKIVESTESQLOILMESRSKYDAVSKFEKSSSSSSSDS 120
QY 158 STKPEASTAKPNKPTPEGEKVAEAKKVEAEKAKQKEDRRNYPITTYKTLELEIA 216
DB 121 STKPEASTAKPNKPTPEGEKVAEAKKVEAEKAKQKEDRRNYPITTYKTLELEIA 180
QY 217 ESDVEVKKAELVYKVAKEPRDEQIKQAEAEVESKQAEATRLKKITDREAEAEAKR 276
DB 181 ESDVEVKKAELVYKVAKEPRDEQIKQAEAEVESKQAEATRLKKITDREAEAEAKR 240
QY 277 RADAKEQKPKGRARAGVPGELATPDKKENDAKSSDSSVGEEITLPSPLKPEKKVAEAK 336
DB 241 RADAKEQKPKGRARAGVPGELATPDKKENDAKSSDSSVGEEITLPSPLKPEKKVAEAK 300
QY 337 KYEAKKKAADQKEDRRNYPITTYKTLELEIAESDVEVKKAELVYKBEAKPEPNEKY 396
DB 301 KYEAKKKAADQKEDRRNYPITTYKTLELEIAESDVEVKKAELVYKBEAKPEPNEKY 360
QY 397 KOAKAEVESKKAETRLKIKTDKKAEBEAKRKAEDKVEKEPAEOPQAPAKAEKP 456
DB 361 KOAKAEVESKKAETRLKIKTDKKAEBEAKRKAEDKVEKEPAEOPQAPAKAEKP 420
QY 457 APAPKPEPAEQPKAEKPADQAEEDVYARSEEBYNRLTQOQPKTEKPAQBPSTP 511
DB 421 APAPKPEPAEQPKAEKPADQAEEDVYARSEEBYNRLTQOQPKTEKPAQBPSTP 475

RESULT 5
AA192099
ID AA192099 standard; Protein; 655 AA.
XX

AC AA192099;
XX
XX 01-FEB-2000 (first entry)
DT
XX Choline binding protein A (Cbpa).
DE
XX Choline binding protein; Cbpa; adhesin; immunogen;
KW vaccine; diagnosis; therapy; pneumococcus; sepsis; meningitis;
KW otitis media; pneumonia.
XX
XX Streptococcus pneumoniae.
OS
XX
XX WO951188-A2.
PN
XX
XX 14-OCT-1999.
PD
XX
XX 07-APR-1999; 99WO-US07669.
PF
XX
XX 07-APR-1998; 98US-0056019.
PR
XX 07-APR-1998; 98US-0080878.
PR
XX
XX (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
PA (MEDI-) MEDIMUNE INC.
XX
XX Tuomanen EI, Maure HR, Wizemann TM, Johnson LS, Koenig S;
PI WPI, 1999-633690/54.
XX N-PSDB; AA234452.
DR
XX
XX New N-terminal choline binding protein A truncate polypeptides, used to
PT develop products for the diagnosis, prevention and treatment of
PT pneumococcal infections -
XX
XX Disclosure; Page 88-93; 160pp; English.
PS
XX
XX This sequence represents the choline binding protein A (Cbpa)
CC of Streptococcus pneumoniae serotype type 4. The invention provides
CC novel N-terminal Cbpa truncated polypeptides (see AA192098-110 and
CC AA192179-90) and polynucleotides encoding them, host-vector systems,
CC and antibodies that specifically bind to the polypeptides. The
CC invention also relates to vaccines including the polypeptides,
CC which provide protection or elicit protective antibodies to
CC bacterial infection, specifically pneumococcus, and to antibodies
CC and antagonists against such polypeptides for use in diagnosis and
CC passive immunotherapy. The polypeptides and/or polynucleotides are
CC also useful as competitive inhibitors of bacterial adhesion of
CC pneumococcus.
XX
XX Sequence 655 AA;
SQ

Query Match 92.6%; Score 2350.5; DB 20; Length 655;
Best Local Similarity 99.8%; Pred. No. 1.2e-141;

Matches 474; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 38 ENEGATQVPTSSNRANESQAQEQOPKLDSEBDKARKEVEEYVKIIGESYAKSTKKRH 97
DB 1 ENEGATQVPTSSNRANESQAQEQOPKLDSEBDKARKEVEEYVKIIGESYAKSTKKRH 60
QY 98 TITVALVNLNINKNEYLINKIVESTESQLOILMESRSKYDAVSKFEKSSSSSSSDS 157
DB 61 TITVALVNLNINKNEYLINKIVESTESQLOILMESRSKYDAVSKFEKSSSSSSSDS 120
QY 158 STKPEASTAKPNKPTPEGEKVAEAKKVEAEKAKQKEDRRNYPITTYKTLELEIA 216
DB 121 STKPEASTAKPNKPTPEGEKVAEAKKVEAEKAKQKEDRRNYPITTYKTLELEIA 180
QY 217 ESDVEVKKAELVYKVAKEPRDEQIKQAEAEVESKQAEATRLKKITDREAEAEAKR 276
DB 181 ESDVEVKKAELVYKVAKEPRDEQIKQAEAEVESKQAEATRLKKITDREAEAEAKR 240
QY 277 RADAKEQKPKGRARAGVPGELATPDKKENDAKSSDSSVGEEITLPSPLKPEKKVAEAK 336
DB 241 RADAKEQKPKGRARAGVPGELATPDKKENDAKSSDSSVGEEITLPSPLKPEKKVAEAK 300

QY 337 KVEAKKKAEDEDRNYPNTYKTLLEIAESDVEVKAELELVKEAEKPEPNEKV 396
DB 301 KVEAKKKAEDEDRNYPNTYKTLLEIAESDVEVKAELELVKEAEKPEPNEKV 360
QY 397 KQAKAEVSKKAEATRLKIKTDKKAEEBAKKAEDKVEKPAEOPAPAPAKAEKP 456
DB 361 KQAKAEVSKKAEATRLKIKTDKKAEEBAKKAEDKVEKPAEOPAPAPAKAEKP 420
QY 457 APAPPEPAPAEOPKAEKPADQAEEDYARSSSEYNRLTQOOPPTTEKPAOPSTP 511
DB 421 APAPPEPAPAEOPKAEKPADQAEEDYARSSSEYNRLTQOOPPTTEKPAOPSTP 475

RESULT 6
AA49250
ID AA49250 standard; Protein; 460 AA.
AC AA49250;
XX 07-FEB-2000 (first entry)
DT N-terminal region of Cbpa polypeptide Ntype4.
DE Choline binding protein A; Cbpa; truncate; immune response; infection;
KW pneumococcal bacterium; vaccine.
XX Streptococcus pneumoniae.
OS
XX MO9951187-A2.
XX 14-OCT-1999.
XX 07-APR-1999; 99WO-US07668.
XX 07-APR-1998; 98US-0056019.
PR 07-APR-1998; 98US-0080878.
XX
XX (STUD-) ST JUDE CHILDREN'S RES HOSPITAL.
XX
XX Tuomanen EI, Masure HR;
PI WPI; 1999-620161/53.
DR Novel polypeptides, used to develop products for the diagnosis,
PT prevention and treatment of pneumococcal infections
XX
XX Disclosure; Fig 2A-B; 85pp; English.
XX
XX The invention provides novel isolated polypeptides comprising the amino
CC acid sequence of an N-terminal choline binding protein A (Cbpa) truncate
CC (CbpaT). The polypeptides can be selected from sequences shown in
CC AA49225, AA49227, AA49230, AA49231, AA49233 and AA49238. The
CC polypeptides can be used for inducing an immune response in a subject
CC which has been exposed to or infected with a pneumococcal bacterium. They
CC can also be used for preventing infection by a pneumococcal bacterium.
CC Vaccines comprising the polypeptides or encoding nucleic acids can be
CC used for treating a subject infected with or exposed to a pneumococcal
CC bacterium. Antibodies specifically binding the polypeptides can be used
CC for detection and diagnosis and for preventing pneumococcal attachment to
CC a mucosal surface. The products can be used in humans and other animals
CC such as domestic animals, such as feline or canine subjects, farm animals
CC such as bovine, equine, caprine, ovine, and porcine subjects, wild
CC animals (whether in the wild or in a zoological garden), research
CC animals, such as mice, rats, rabbits, goats, sheep, pigs, dogs, cats,
CC i.e. for veterinary medical use.
XX
XX Sequence 460 AA;
SQ

Query Match 88.8%; Score 2252.5; DB 20; Length 460;
Best Local Similarity 99.6%; Pred. No. 1.4e-135;
Matches 457; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

QY 24 VVASLWGSVVHATENEGATVPTSSNRANESQAEQEGOPKLDSEDPKAREVEVYKK 83
DB 2 IVASLWGSVVHATENEGATVPTSSNRANESQAEQEGOPKLDSEDPKAREVEVYKK 61
QY 84 IVGSYAKSTKKRHTITVALVNELNINKNEYLNKIVESTSSQOLIMMESRSKYDEAVS 143
DB 62 IVGSYAKSTKKRHTITVALVNELNINKNEYLNKIVESTSSQOLIMMESRSKYDEAVS 121
QY 144 KFEKSSSSSSSDSTKPEASDTAKPNKPTPEGKVAAKKKVEAEKKAADQKEEDRN 202
DB 122 KFEKSSSSSSSDSTKPEASDTAKPNKPTPEGKVAAKKKVEAEKKAADQKEEDRN 181
QY 203 YPTITTYKTLLEIAESDVEVKAELELVKVAANEPRDQKIKQAEVSKQAEATRLKK 262
DB 182 YPTITTYKTLLEIAESDVEVKAELELVKVAANEPRDQKIKQAEVSKQAEATRLKK 241
QY 263 IKTDREAEAEKARADAKAEQCKPKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPS 322
DB 242 IKTDREAEAEKARADAKAEQCKPKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPS 301
QY 323 PSLKPEKKVAAEKKVEAEKKAEDQKEEDRNYPNTYKTLLEIAESDVEVKAELEL 382
DB 302 PSLKPEKKVAAEKKVEAEKKAEDQKEEDRNYPNTYKTLLEIAESDVEVKAELEL 361
QY 383 VKEAEKPEPNEKVQAKAEVSKKAEATRLKIKTDKKAEEBAKKAEDKVEKPA 442
DB 362 VKEAEKPEPNEKVQAKAEVSKKAEATRLKIKTDKKAEEBAKKAEDKVEKPA 421
QY 443 EQPAPAPAKAEKPAPKPPENPAEQPAEKPADQAE 481
DB 422 EQPAPAPAKAEKPAPKPPENPAEQPAEKPADQAE 460

RESULT 7
AA492189
ID AA492189 standard; Protein; 460 AA.
AC AA492189;
XX 01-FEB-2000 (first entry)
DT N-terminal choline binding protein A (Cbpa) truncate.
DE Choline binding protein; Cbpa; truncate; adhesin; immunogen;
KW vaccine; diagnosis; therapy; pneumococcus; sepsis; meningitis;
XX otitis media; pneumonia.
XX Streptococcus pneumoniae.
OS Synthetic.
XX
XX MO9951188-A2.
XX 14-OCT-1999.
XX 07-APR-1999; 99WO-US07669.
XX 07-APR-1998; 98US-0056019.
PR 07-APR-1998; 98US-0080878.
XX
XX (STUD-) ST JUDE CHILDREN'S RES HOSPITAL.
XX (MEDI-) MEDIMUNE INC.
XX Tuomanen EI, Masure HR, Wizemann TW, Johnson LS, Koenig S;
PI WPI; 1999-633690/54.
DR New N-terminal choline binding protein A truncate polypeptides, used to
PT develop products for the diagnosis, prevention and treatment of
PT pneumococcal infections
XX
XX Claim 47; Fig 2A-B; 160pp; English.
XX The present sequence represents an N-terminal choline binding

CC protein A (Cbpa) truncate, denoted Ntype 4 Cbpa trunc, derived from
CC Streptococcus pneumoniae Cbpa. Claimed vaccines contain and
CC N-terminal Cbpa truncate such as the present sequence, or a
CC polypeptide comprising a conserved region of the Cbpa truncate.
CC The vaccines provide protection or elicit protective antibodies to
CC bacterial infection, specifically pneumococcus. Antibodies and
CC and antagonists against the N-terminal Cbpa truncates are used in
CC diagnosis and passive immunotherapy.

XX Sequence 460 AA;

Query Match 88.7%; Score 2249.5; DB 20; Length 460;
Best Local Similarity 99.3%; Pred. No. 2.1e-135;
Matches 456; Conservative 2; Mismatches 0; Indels 1; Gaps 1;

QY 24 VVASLWMSVVAHATENEGATQVPTSSNRANESQAEQEGQPKLDSERDKARKEVEEYVK 83
DB :|||||
2 IVASLWMSVVAHATENEGATQVPTSSNRANESQAEQEGQPKLDSERDKARKEVEEYVK 61
QY IVGESYAKSTKKRHTTVALVNLNINIKVEYLNKIVESTSESQOILMMESRSKYDEAVS 143
DB :|||||
62 IVGESYAKSTKKRHTTVALVNLNINIKVEYLNKIVESTSESQOILMMESRSKYDEAVS 121
QY KFEKSSSSSSSDSTKPEASDTAKPNKPTPEGEKVAEAKKVEBA-KKAKDQKEEDRN 202
DB 122 KFEKSSSSSSSDSTKPEASDTAKPNKPTPEGEKVAEAKKVEBAKAKDQKEEDRN 181
QY 203 YPTITYKTLELEIASDVEVKAELELVKVKANPEPDEOKIKQAEAVESKQAEATRLKK 262
DB 182 YPTITYKTLELEIASDVEVKAELELVKVKANPEPDEOKIKQAEAVESKQAEATRLKK 241
QY 263 IKTDEEAEAEAKRADAKQEGKPKGRAGVPGELATPDKKENDAKSSDSVGEETLPS 322
DB 242 IKTDEEAEAEAKRADAKQEGKPKGRAGVPGELATPDKKENDAKSSDSVGEETLPS 301
QY 323 PSLKPEKKVAEAEKVEAKKAEADQKEEDRNYPNTYKTLELEIASDVEVKAELEL 382
DB 302 PSLKPEKKVAEAEKVEAKKAEADQKEEDRNYPNTYKTLELEIASDVEVKAELEL 361
QY 383 VKEAKEPRNEEKVQAKAEVESKKAATRLKIKTRDKKAEAEAKKRAEEDKYKEKPA 442
DB 362 VKEAKEPRNEEKVQAKAEVESKKAATRLKIKTRDKKAEAEAKKRAEEDKYKEKPA 421
QY 443 EQOPAPAPKAKPAPKPPENPAEQPAEKPADQAE 481
DB 422 EQOPAPAPKAKPAPKPPENPAEQPAEKPADQAE 460

RESULT 8

AAV49251
ID AAV49251 standard; Protein; 459 AA.

XX AAV49251;

DT 07-FEB-2000 (first entry)

DE N-terminal region of Cbpa polypeptide ATCC4.

KM Choline binding protein A; Cbpa; truncate; immune response; infection;
XX pneumococcal bacterium; vaccine.

OS Streptococcus pneumoniae.

PN W09951187-A2.

PD 14-OCT-1999.

PF 07-APR-1999; 99MO-US07668.

PR 07-APR-1998; 98US-0056019.

PR 07-APR-1998; 98US-0080878.

PA (SUD-) ST JUDE CHILDREN'S RES HOSPITAL.

XX Tuomanen EI, Maure HR.

DR WPI; 1999-620161/53.

PT Novel polypeptides, used to develop products for the diagnosis,
XX prevention and treatment of pneumococcal infections -

PS Disclosure: Fig 2A-B; 85pp; English.

CC The invention provides novel isolated polypeptides comprising the amino
CC acid sequence of an N-terminal choline binding protein A (Cbpa) truncate
CC (CbpaT). The polypeptides can be selected from sequences shown in
CC AAV49225, AAV49227, AAV49230, AAV49231, AAV49233 and AAV49238. The
CC polypeptides can be used for inducing an immune response in a subject
CC which has been exposed to or infected with a pneumococcal bacterium.
CC can also be used for preventing infection by a pneumococcal bacterium.
CC Vaccines comprising the polypeptides or encoding nucleic acids can be
CC used for treating a subject infected with or exposed to a pneumococcal
CC bacterium. Antibodies specifically binding the polypeptides can be used
CC for detection and diagnosis and for preventing pneumococcal attachment to
CC a mucosal surface. The products can be used in humans and other animals
CC such as domestic animals, such as feline or canine subjects, farm animals
CC such as bovine, equine, caprine, ovine, and porcine subjects, wild
CC animals (whether in the wild or in a zoological garden), research
CC animals, such as mice, rats, rabbits, goats, sheep, pigs, dogs, cats,
CC i.e. for veterinary medical use.

XX Sequence 459 AA;

Query Match 88.1%; Score 2235.5; DB 20; Length 459;
Best Local Similarity 98.7%; Pred. No. 1.7e-114;
Matches 453; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 24 VVASLWMSVVAHATENEGATQVPTSSNRANESQAEQEGQPKLDSERDKARKEVEEYVK 83
DB 1 IVASLWMSVVAHATENEGATQVPTSSNRANESQAEQEGQPKLDSERDKARKEVEEYVK 60
QY 84 IVGESYAKSTKKRHTTVALVNLNINIKVEYLNKIVESTSESQOILMMESRSKYDEAVS 143
DB 61 IVGESYAKSTKKRHTTVALVNLNINIKVEYLNKIVESTSESQOILMMESRSKYDEAVS 120
QY 144 KFEKSSSSSSSDSTKPEASDTAKPNKPTPEGEKVAEAKKVEBA-KKAKDQKEEDRN 202
DB 121 KFEKSSSSSSSDSTKPEASDTAKPNKPTPEGEKVAEAKKVEBAKAKDQKEEDRN 180
QY 203 YPTITYKTLELEIASDVEVKAELELVKVKANPEPDEOKIKQAEAVESKQAEATRLKK 262
DB 181 YPTITYKTLELEIASDVEVKAELELVKVKANPEPDEOKIKQAEAVESKQAEATRLKK 240
QY 263 IKTDEEAEAEAKRADAKQEGKPKGRAGVPGELATPDKKENDAKSSDSVGEETLPS 322
DB 241 IKTDEEAEAEAKRADAKQEGKPKGRAGVPGELATPDKKENDAKSSDSVGEETLPS 300
QY 323 PSLKPEKKVAEAEKVEAKKAEADQKEEDRNYPNTYKTLELEIASDVEVKAELEL 382
DB 301 PSLKPEKKVAEAEKVEAKKAEADQKEEDRNYPNTYKTLELEIASDVEVKAELEL 360
QY 383 VKEAKEPRNEEKVQAKAEVESKKAATRLKIKTRDKKAEAEAKKRAEEDKYKEKPA 442
DB 361 VKEAKEPRNEEKVQAKAEVESKKAATRLKIKTRDKKAEAEAKKRAEEDKYKEKPA 420
QY 443 EQOPAPAPKAKPAPKPPENPAEQPAEKPADQAE 481
DB 421 EQOPAPAPKAKPAPKPPENPAEQPAEKPADQAE 459

RESULT 9

AAV32190
ID AAV32190 standard; Protein; 459 AA.

XX AAV32190;

DT	01-FEB-2000	(first entry)
XX	N-terminal choline binding protein A (Cbpa) truncate.	
DE		
XX	Choline binding protein; CbpA; truncate; adhesin; immunogen;	
KW	vaccine; diagnosis; therapy: pneumococcus; sepsis; meningitis;	
KW	otitis media; pneumonia.	
XX		
OS	Streptococcus pneumoniae.	
OS	Synthetic.	
XX		
PV	MO9951188-AZ.	
PD		
PF	14-OCT-1999.	
XX		
PF	07-APR-1999; 99WO-US07669.	
XX		
PR	07-APR-1998; 98US-0056019.	
XX		
PA	07-APR-1998; 98US-0080876.	
PA		
PA	(STUD-) ST JUDE CHILDREN'S RES HOSPITAL.	
PA	(MEDI-) MEDIMUNE INC.	
PI		
XX	Tuomanen EI, Maure HR, Wizemann TM, Johnson LS, Koenig S;	
DR	WPI, 1999-633690/54.	
PT		
PT	New N-terminal choline binding protein A truncate polypeptides, used to	
PT	develop products for the diagnosis, prevention and treatment of	
PS	pneumococcal infections -	
XX		
PS	Claim 47; Fig 2A-B; 160pp; English.	
CC		
CC	The present sequence represents an N-terminal choline binding	
CC	protein A (Cbpa) truncate, denoted ATCC4 Cbpa trun, derived from	
CC	Streptococcus pneumoniae Cbpa. Claimed vaccines contain and	
CC	N-terminal Cbpa truncate such as the present sequence, or a	
CC	polypeptide comprising a conserved region of the Cbpa truncate.	
CC	The vaccines provide protection or elicit protective antibodies to	
CC	bacterial infection, specifically pneumococcus. Antibodies and	
CC	antagonists against the N-terminal Cbpa truncates are used in	
CC	diagnosis and passive immunotherapy.	
CC		
SO	Sequence 459 AA;	
Query Match	88.1%; Score 2235.5; .DB 20; Length 459;	
Best Local Similarity	98.7%; Pred. No.1.7e-134;	
Matches 453; Conservative	2; Mismatches 3; Indels 1; Gaps 1,	
OY	24 VVASLVMSVVHATNEGATOVPPSSNNANESQAEQGQPKLDSERKAREVEEYVK 83	
Db	1 IVASLVMSVVHATENEGATOVPPSSNPNANESQAEGQPKLDSERKAREVEEYVK 60	
OY	84 IVGESYAKSTKKRHITITVALVNLNINNEYLNKIIVESTSESOLQILMMESRSKYDAVS 143	
Db	61 IVGSYASTKKRRHITITVALVNLNININEYLNKIIVESTSESOLQILMMESRSKYDAVS 120	
OY	144 KFEKSDSSSSSDSSTKEPEASDTAKPNPTEPEGKVAEAKKVEEA-KYAKDQEEDRRN 202	
Db	121 KFEKSDSSSSSDSSTKEPEASDTAKPNPTEPEGKVAEAKKVEEVEEKAKDQKEEDRN 180	
OY	203 YPLTTYKTLEIEIASDVEYKKAEIELVKVANPERDCKIKQAEAFVESQKAETRLKX 262	
Db	181 YPLTTYKTLEIEIASDVEYKKAEIELVKVANPERDQKIKQAEAFVESQKAETRLKX 240	
OY	263 IKTDREAEAEBAKRADAKEQKPGRAKRGVPGLATLPDKKENDAKSDSVGSETLPS 322	
Db	241 IKTDREAEAEBAKRADAKEQKPGRGRGVGELATLPDKKENDAKSDSVGSETLPS 300	
OY	323 PSLPDEKKVAAEAKKVEAKKAABEQKEDERRNPNTNYKTLEIEIASDVEYKKAEIEL 382	
Db	301 PSLPDEKKVAAEAKKVEAKKAABEQKEDERRNPNTNYKTLEIEIASDVEYKKAEIEL 360	

Qy		368	VVEEAKPEPNEKRVQAARAEVSSSKAERLEKITDRCKAAEEBAKRKAABEPKYEKRA	442
Db		361	VVEEAKEPEPNEKRVQAAREVSSSKAAEARLRKITDRCKAAEEBAKRKAABEDKVAKERA	420
Qy		443	EOPAPAPAKKPPAPAPKPENPAOPKAKEPADOOAE	481
Db		421	EOPAPAPAKTKPPAPAPKPENPAOPKAKEPADOOAE	459
 RESULT 10 AAV49143 ID AAV49143 standard; Protein; 446 AA. XX AC AAV49143; XX DT 17-JAN-2000 (first entry) XX DE Amino acid sequence of choline-binding protein fragment #7. XX KW Truncated surface binding protein; alpha helix; choline binding protein; KM vaccinia; invasive bacterial infection; otitis media; sepsis; KM meningitis; lobar pneumonia infection; antibody; immature immune system; KW immunocompromised. XX OS Streptococcus pneumoniae. XX PN WO951266-A2. XX PD 14-OCT-1999. XX PF 06-APR-1999; 99WO-US07680. XX PR 07-APR-1998; 98US-0080878. PR 15-MAY-1998; 98US-0085743. XX PA (MEDI-) MEDIMUNE INC. XX PI Wizemann TM, Koenig S, Johnson LS; XX DR WPI; 1999-601465/51. XX DR N-PSDB; AAZ31083. XX PT New pneumococcal proteins useful as vaccines and for diagnosis of PT pneumococcal infections - PS Claim 10; Page 72-73; 98pp; English.				
Qy	Sequence	446 AA;		
Query Match	86.5%;	Score 2194.5;	DB 20;	Length 446;
Best Local Similarity	99.8%;	Pred. No. 6.5e-132;		
Matches 445;	Conservative	0;	Mismatches	0;
			Indels	1;
			Gaps	1

```
QY 37 TENEGATOVPTSSNRANESQAEQGPCKLDSERDKARKEVEYVKTIIGESYAKSTKR 96
DB 1 TENEGATOVPTSSNRANESQAEQGPCKLDSERDKARKEVEYVKTIIGESYAKSTKR 60
QY 97 HTITVALVNLINIKNEYLINKIVESTSESQIILMESRSKYDEAVSKFEKSSSSSSD 156
DB 61 HTITVALVNLINIKNEYLINKIVESTSESQIILMESRSKYDEAVSKFEKSSSSSSD 120
QY 157 SSTKEPADTAKPNKPTPEGEKVAEAKKVEEA-KKADQKEEDRRNPPTITTYKTLEI 215
DB 121 SSTKEASDTAKPNKPTPEGEKVAEAKKVEEA-KKADQKEEDRRNPPTITTYKTLEI 180
QY 216 AESDVEVKKAELELVKVKANEPDEOKIKQAEAVESKQAEATRLKIKITDREAEAEAK 275
DB 181 AESDVEVKKAELELVKVKANEPDEOKIKQAEAVESKQAEATRLKIKITDREAEAEAK 240
QY 276 RRAADKEQGPGRKRGVPGELATPDKKENDAKSSDSSVGEETLPSPSLKEKKVAEAE 335
DB 241 RRAADKEQGPGRKRGVPGELATPDKKENDAKSSDSSVGEETLPSPSLKEKKVAEAE 300
QY 336 KKVBEAKKKAEDQKEEDRRNPPTITTYKTLEIASEDVVKKAELELVKEAEKPRNEEK 395
DB 301 KKVBEAKKKAEDQKEEDRRNPPTITTYKTLEIASEDVVKKAELELVKEAEKPRNEEK 360
QY 336 VQAKAEVSKKAEATRLKIKITDRKKAEEAKKRAAEEDVKKEKPAQOPAPAPKAEK 455
DB 361 VQAKAEVSKKAEATRLKIKITDRKKAEEAKKRAAEEDVKKEKPAQOPAPAPKAEK 420
QY 456 PAPAPKPNPAPQPAEKPADQAE 481
DB 421 PAPAPKPNPAPQPAEKPADQAE 446
RESULT 11
AA49140
AA49140 standard; Protein; 446 AA.
AC AA49140;
DT 17-JAN-2000 (first entry)
DE Amino acid sequence of choline-binding protein fragment #4.
KW Truncated surface binding protein; alpha helix; choline binding protein;
KW vaccine; invasive bacterial infection; otitis media; sepsis;
KW meningitis; lobar pneumonia infection; antibody; immature immune system;
KW immunocompromised.
OS Streptococcus pneumoniae.
XX
XX Key Location/Qualifiers
XX FH Misc-difference /note= "Encoded by AAT"
XX FT Misc-difference /note= "Encoded by TAC"
XX FT Misc-difference /note= "Encoded by TTA"
XX FT Misc-difference /note= "Encoded by TTA"
XX FT Misc-difference /note= "Encoded by CAA"
XX FT Misc-difference /note= "Encoded by CAA"
XX PN MO9951266-A2.
XX PD 14-OCT-1999.
XX PF 06-APR-1999; 99WO-US07680.
XX PR 07-APR-1998; 98US-0080878.
XX PR 15-MAY-1998; 98US-0085743.
XX PA (MEDI-) MEDIMUNE INC.
XX PI Wizemann TM, Koenig S, Johnson LS;
```

```
XX WPI: 1999-601465/51.
DR N-PSDB; AA231080.
XX
PT New pneumococcal proteins useful as vaccines and for diagnosis of
DB pneumococcal infections -
XX Claim 10; Page 68-69; 98pp; English.
XX
XX AA49137-49152 are amino acid sequences that are fragments of choline
CC binding proteins (CBP). The fragments of the protein are the alpha helix
CC forming parts of the CBPs from Streptococcus pneumoniae. The
CC polypeptides do not contain the actual choline binding fragment. The
CC polypeptides and the nucleotide sequences that encode them
CC (AA231077-231092) are used in the invention, which relates to polypeptide
CC truncates of a pneumococcal surface binding protein containing the highly
CC conserved immunogenic alpha helical portion and no choline binding
CC portion. The polypeptides are used as immunogens in a bacterial vaccine.
CC The vaccine can be used for preventing (immunizing) or treating invasive
CC bacterial (especially pneumococcal) infections, especially otitis media
CC (caused by S.pneumoniae), sepsis, meningitis and lobar pneumonia
CC infections. Antibodies raised against the polypeptide are useful for
CC detection, prevention (passive immunity) and treatment of S. pneumoniae
CC infections. The vaccines are especially useful in immunocompromised
CC patients, those with an immature immune system, or patients with an on
CC going pneumococcal infection. The vaccine avoids unnecessary expense and
CC provides broad protection against a range of pneumococcal serotypes and
CC it produces an improved and enhanced effect in preventing bacterial
CC infections.
SQ Sequence 446 AA;
```

Query Match 85.8%; Score 2177.5; DB 20; Length 446;

Best Local Similarity 98.9%; Pred. No. 7.9e-131;

Matches 441; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

```
QY 37 TENEGATOVPTSSNRANESQAEQGPCKLDSERDKARKEVEYVKTIIGESYAKSTKR 96
DB 1 TENEGATOVPTSSNRANESQAEQGPCKLDSERDKARKEVEYVKTIIGESYAKSTKR 60
QY 97 HTITVALVNLINIKNEYLINKIVESTSESQIILMESRSKYDEAVSKFEKSSSSSSD 156
DB 61 HTITVALVNLINIKNEYLINKIVESTSESQIILMESRSKYDEAVSKFEKSSSSSSD 120
QY 157 SSTKEPADTAKPNKPTPEGEKVAEAKKVEEA-KKADQKEEDRRNPPTITTYKTLEI 215
DB 121 SSTKEASDTAKPNKPTPEGEKVAEAKKVEEA-KKADQKEEDRRNPPTITTYKTLEI 180
QY 216 AESDVEVKKAELELVKVKANEPDEOKIKQAEAVESKQAEATRLKIKITDREAEAEAK 275
DB 181 AESDVEVKKAELELVKVKANEPDEOKIKQAEAVESKQAEATRLKIKITDREAEAEAK 240
QY 276 RRAADKEQGPGRKRGVPGELATPDKKENDAKSSDSSVGEETLPSPSLKEKKVAEAE 335
DB 241 RRAADKEQGPGRKRGVPGELATPDKKENDAKSSDSSVGEETLPSPSLKEKKVAEAE 300
QY 336 KKVBEAKKKAEDQKEEDRRNPPTITTYKTLEIASEDVVKKAELELVKEAEKPRNEEK 395
DB 301 KKVBEAKKKAEDQKEEDRRNPPTITTYKTLEIASEDVVKKAELELVKEAEKPRNEEK 360
QY 336 VQAKAEVSKKAEATRLKIKITDRKKAEEAKKRAAEEDVKKEKPAQOPAPAPKAEK 455
DB 361 VQAKAEVSKKAEATRLKIKITDRKKAEEAKKRAAEEDVKKEKPAQOPAPAPKAEK 420
QY 456 PAPAPKPNPAPQPAEKPADQAE 481
DB 421 PAPAPKPNPAPQPAEKPADQAE 446
RESULT 12
AA49238
AA49238 standard; Protein; 428 AA.
```

AC AAY49238;
 XX
 DT 07-FEB-2000 (first entry)
 XX
 DE Choline binding protein A (Cbpa) serotype type 4 amino acid sequence.
 XX
 KM Choline binding protein A; Cbpa; truncate; immune response; infection;
 KW pneumococcal bacterium; vaccine.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN MO9951187-A2.
 XX
 PD 14-OCT-1999.
 XX
 PF 07-APR-1999; 99WO-US07668.
 XX
 PR 07-APR-1998; 98US-0056019.
 PR 07-APR-1998; 98US-0080878.
 XX
 PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
 PI Tuomanen EI, Masure HR;
 XX
 DR WPI; 1999-620161/53.
 XX
 PT Novel polypeptides, used to develop products for the diagnosis,
 PT prevention and treatment of pneumococcal infections
 XX
 PS Claim 7; Page 7; 85pp; English.
 XX
 CC The invention provides novel isolated polypeptides comprising the amino
 CC acid sequence of an N-terminal choline binding protein A (Cbpa) truncate
 CC (CBPAT). The polypeptides can be selected from sequences shown in
 CC AAY49235, AAY49237, AAY49230, AAY49231, AAY49233 and AAY49238. The
 CC polypeptides can be used for inducing an immune response in a subject
 CC which has been exposed to or infected with a pneumococcal bacterium. They
 CC can also be used for preventing infection by a pneumococcal bacterium.
 CC Vaccines comprising the polypeptides or encoding nucleic acids can be
 CC used for treating a subject infected with or exposed to a pneumococcal
 CC bacterium. Antibodies specifically binding the polypeptides can be used
 CC for detection and diagnosis and for preventing pneumococcal attachment to
 CC a mucosal surface. The products can be used in humans and other animals
 CC such as domestic animals, such as feline or canine subjects, farm animals
 CC such as bovine, equine, caprine, ovine, and porcine subjects, wild
 CC animals (whether in the wild or in a zoological garden), research
 CC animals, such as mice, rats, rabbits, goats, sheep, pigs, dogs, cats,
 CC i.e. for veterinary medical use.
 XX
 SQ Sequence 428 AA;
 Query Match 82.8%; Score 2101.5; DB 20; Length 428;
 Best Local Similarity 99.8%; Pred. No. 5.2e-126;
 Matches 427; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 QY 38 ENEGATQVPTSSNANESQAOEGOPKTLDESRDARKEVEYKTVGSEYAKSTKRH 97
 DB 1 ENEGATQVPTSSNANESQAOEGOPKTLDESRDARKEVEYKTVGSEYAKSTKRH 60
 QY 98 TITVALVNLNNIKNEYLNKIVESTESQOILMMESRSKDEAVSKPEKSSSSSSSDS 157
 DB 61 TITVALVNLNNIKNEYLNKIVESTESQOILMMESRSKDEAVSKPEKSSSSSSSDS 120
 QY 158 STKEPASTAKPKPTPEGEVVAEAKKVEBA-KKAKDQKEEDRRNPTTYKTLELEIA 216
 DB 121 STKEPASTAKPKPTPEGEVVAEAKKVEBA-KKAKDQKEEDRRNPTTYKTLELEIA 180
 QY 217 ESDVEVKKAELVKKVANEPRDOKIKQAEVYSKQAEKTRTKTKTDEEPAEERAKR 276
 DB 181 ESDVEVKKAELVKKVANEPRDOKIKQAEVYSKQAEKTRTKTKTDEEPAEERAKR 240
 QY 277 RADAKEOGKPKRAKRGVPGELATPDKKENDAKSSDSSVGSEETLPSPSLKPEKKVAEAK 336
 DB 277 RADAKEOGKPKRAKRGVPGELATPDKKENDAKSSDSSVGSEETLPSPSLKPEKKVAEAK 336

DB 241 RADAKEOGKPKRAKRGVPGELATPDKKENDAKSSDSSVGSEETLPSPSLKPEKKVAEAK 300
 QY 337 KVEEAKKKAEDQEKEDERNYNTNTYKTLELEIAESDVEVKKAELVVEEAKPEERNEKV 396
 DB 301 KVEEAKKKAEDQEKEDERNYNTNTYKTLELEIASDVEVKKAELVVEEAKPEERNEKV 360
 QY 397 KQAEVYSKKAELTRLEKIKTDKKAEEBAKRAEEDKVEKPAEQOPAPAPAKAEK 456
 DB 361 KQAEVYSKKAELTRLEKIKTDKKAEEBAKRAEEDKVEKPAEQOPAPAPAKAEK 420
 QY 457 APAPKPEK 464
 DB 421 APAPKPEK 428
 RESULT 13
 AAY32110
 ID AAY32110 standard; Protein; 428 AA.
 XX
 AC AAY32110;
 XX
 DT 01-FEB-2000 (first entry)
 XX
 DE Choline binding protein A (Cbpa) N-terminal truncate.
 XX
 KM Choline binding protein; Cbpa; truncate; adhesion; immunogen;
 KM vaccine; diagnosis; therapy; pneumococcus; sepsis; meningitis;
 KM otitis media; pneumonia.
 XX
 OS Streptococcus pneumoniae.
 OS Synthetic.
 XX
 PN MO9951188-A2.
 XX
 PD 14-OCT-1999.
 XX
 PF 07-APR-1999; 99WO-US07669.
 XX
 PR 07-APR-1998; 98US-0056019.
 PR 07-APR-1998; 98US-0080878.
 XX
 PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
 PA (MEDI-) MEDIMUNE INC.
 PI Tuomanen EI, Masure HR, Witzmann TM, Johnson LS, Koenig S;
 XX
 DR WPI; 1999-633690/54.
 XX
 PT New N-terminal choline binding protein A truncate polypeptides, used to
 PT develop products for the diagnosis, prevention and treatment of
 PT pneumococcal infections
 XX
 PS Claim 7; Page 119-122; 160pp; English.
 XX
 CC The present sequence represents a new N-terminal choline binding
 CC protein A (Cbpa) truncate that maintains the tertiary structure of
 CC Streptococcus pneumoniae serotype type 4 Cbpa native protein.
 CC Host-vector systems for production of N-terminal Cbpa truncates are
 CC provided. The invention relates generally to novel N-terminal
 CC Cbpa truncates and polynucleotides encoding them, host-vector
 CC systems, and antibodies that specifically bind to the truncates.
 CC The invention also relates to vaccines including N-terminal Cbpa
 CC truncates, which provide protection or elicit protective antibodies
 CC to bacterial infection, specifically pneumococcus, and to use of
 CC antibodies and antagonists against such polypeptides in diagnosis
 CC and passive immunotherapy. The polypeptides and/or polynucleotides
 CC are also useful as competitive inhibitors of bacterial adhesion of
 CC pneumococcus.
 XX
 SQ Sequence 428 AA;
 Query Match 82.8%; Score 2101.5; DB 20; Length 428;
 Best Local Similarity 99.8%; Pred. No. 5.2e-126;

Matches 427; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 38 ENEGATQVPTSSNRANESQAEQEQPKLDSERDKARKEVEYVKIIGESYAKSTKRH 97
DB 1 ENEGATQVPTSSNRANESQAEQEQPKLDSERDKARKEVEYVKIIGESYAKSTKRH 60
QY 98 TITVALVNLNINIKNEYLINKIVESTSESQOLIMMESRSKYDEAVSKFEKSSSSSSDS 157
DB 61 TITVALVNLNINIKNEYLINKIVESTSESQOLIMMESRSKYDEAVSKFEKSSSSSSDS 120
QY 158 STKPASDTAKPNKPTBEGEKVAEAKKKVEEA-KKAKDQKEDRRNYPITYKTLELEIA 216
DB 121 STKPASDTAKPNKPTBEGEKVAEAKKKVEEA-KKAKDQKEDRRNYPITYKTLELEIA 180
QY 217 ESDVVKKAELVVKVANEPRDEQIKQAEAEVSKQAEATRLKIKITDREAEAEAKR 276
DB 181 ESDVVKKAELVVKVANEPRDEQIKQAEAEVSKQAEATRLKIKITDREAEAEAKR 240
QY 277 RADAKEQCKPKGRAKRGVPGELATPDKKENDAKSSDSVGEETLPSPLKREKVAEAK 336
DB 241 RADAKEQCKPKGRAKRGVPGELATPDKKENDAKSSDSVGEETLPSPLKREKVAEAK 300
QY 337 KVEEAKKKAEDQKEDRRNYPITYKTLELEIASDVVKKAELVKEAEKPRNEEKV 396
DB 301 KVEEAKKKAEDQKEDRRNYPITYKTLELEIASDVVKKAELVKEAEKPRNEEKV 360
QY 397 KQAKAEVSKKAEAATRLKIKITDRKKAEBEAKKAAEBEDKVEKPAEOPQAPAKAEKP 456
DB 361 KQAKAEVSKKAEAATRLKIKITDRKKAEBEAKKAAEBEDKVEKPAEOPQAPAKAEKP 420
QY 457 APAPKREN 464
DB 421 APAPKREN 428

RESULT 14
AAV49225
ID AAV49225 standard; protein; 406 AA.
XX
AC AAV49225;
XX
DT 07-FEB-2000 (first entry)
XX
DE Polypeptide R2 of the N-terminal truncate of Cbpa serotype type 4.
XX
KW Choline binding protein A; Cbpa; truncate; immune response; infection;
KM pneumococcal bacterium; vaccine.
OS Streptococcus pneumoniae.
XX
PN WO9951187-A2.
XX
PD 14-OCT-1999.
XX
PF 07-APR-1999; 99WO-US07668.
XX
PR 07-APR-1998; 98US-0056019.
XX
PR 07-APR-1998; 98US-0080878.
XX
PA (SUD-) ST JUDE CHILDREN'S RES HOSPITAL.
XX
PI Tuomanen EI, Maure HR;
XX
DR WPI; 1999-620161/53.
XX
DR N-PSDB; AA231401.
XX
PT Novel polypeptides, used to develop products for the diagnosis,
XX prevention and treatment of pneumococcal infections
XX
PS Claim 2; Page 7-8; 85pp; English.
XX
CC The invention provides novel isolated polypeptides comprising the amino
CC acid sequence of an N-terminal choline binding protein A (Cbpa) truncate

CC (Cbpa). The polypeptides can be selected from sequences shown in
CC AAV49225, AAV49227, AAV49230, AAV49231, AAV49233 and AAV49238. The
CC polypeptides can be used for inducing an immune response in a subject
CC which has been exposed to or infected with a pneumococcal bacterium. They
CC can also be used for preventing infection by a pneumococcal bacterium.
CC Vaccines comprising the polypeptides or encoding nucleic acids can be
CC used for treating a subject infected with or exposed to a pneumococcal
CC bacterium. Antibodies specifically binding the polypeptides can be used
CC for detection and diagnosis and for preventing pneumococcal attachment to
CC a mucosal surface. The products can be used in humans and other animals
CC such as domestic animals, such as feline or canine subjects, farm animals
CC such as bovine, equine, caprine, ovine, and porcine subjects, wild
CC animals (whether in the wild or in a zoological garden), research
CC animals, such as mice, rats, rabbits, goats, sheep, pigs, dogs, cats,
CC i.e. for veterinary medical use.
XX
SQ Sequence 406 AA;
XX

Query Match 78.0%; Score 1979.5; DB 20; Length 406;
Best Local Similarity 99.8%; Pred. No. 2.9e-118;
Matches 405; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 38 ENEGATQVPTSSNRANESQAEQEQPKLDSERDKARKEVEYVKIIGESYAKSTKRH 97
DB 1 ENEGATQVPTSSNRANESQAEQEQPKLDSERDKARKEVEYVKIIGESYAKSTKRH 60
QY 98 TITVALVNLNINIKNEYLINKIVESTSESQOLIMMESRSKYDEAVSKFEKSSSSSSDS 157
DB 61 TITVALVNLNINIKNEYLINKIVESTSESQOLIMMESRSKYDEAVSKFEKSSSSSSDS 120
QY 158 STKPASDTAKPNKPTBEGEKVAEAKKKVEEA-KKAKDQKEDRRNYPITYKTLELEIA 216
DB 121 STKPASDTAKPNKPTBEGEKVAEAKKKVEEA-KKAKDQKEDRRNYPITYKTLELEIA 180
QY 217 ESDVVKKAELVVKVANEPRDEQIKQAEAEVSKQAEATRLKIKITDREAEAEAKR 276
DB 181 ESDVVKKAELVVKVANEPRDEQIKQAEAEVSKQAEATRLKIKITDREAEAEAKR 240
QY 277 RADAKEQCKPKGRAKRGVPGELATPDKKENDAKSSDSVGEETLPSPLKREKVAEAK 336
DB 241 RADAKEQCKPKGRAKRGVPGELATPDKKENDAKSSDSVGEETLPSPLKREKVAEAK 300
QY 337 KVEEAKKKAEDQKEDRRNYPITYKTLELEIASDVVKKAELVKEAEKPRNEEKV 396
DB 301 KVEEAKKKAEDQKEDRRNYPITYKTLELEIASDVVKKAELVKEAEKPRNEEKV 360
QY 397 KQAKAEVSKKAEAATRLKIKITDRKKAEBEAKKAAEBEDKVEKPA 442
DB 361 KQAKAEVSKKAEAATRLKIKITDRKKAEBEAKKAAEBEDKVEKPA 406

RESULT 15
AAV32098
ID AAV32098 standard; Protein; 406 AA.
XX
AC AAV32098;
XX
DT 01-FEB-2000 (first entry)
XX
DE Choline binding protein A (Cbpa) truncate R2.
XX
KW Choline binding protein; Cbpa; truncate; adhesion; immunogen;
KM vaccine; diagnosis; therapy; pneumococcus; sepsis; meningitis;
OS Streptococcus pneumoniae.
XX
PN WO9951188-A2.
XX
PD 14-OCT-1999.
XX
PR 07-APR-1999; 99WO-US07669.
XX

XX 07-APR-1998: 98US-0056019.
PR 07-APR-1998: 98US-0080878.
XX

PA (STUD-) ST JUDE CHILDREN'S RES HOSPITAL.
PA (MEDI-) MEDIMMUNE INC.
XX

PI Tuomanen EI, Masure HR, Wizemann TM, Johnson LS, Koenig S;
XX WPI: 1999-633690/54.
DR N-PSDB; AAJ34451.
XX

PT New N-terminal choline binding protein A truncate polypeptides, used to
PT develop products for the diagnosis, prevention and treatment of
PT pneumococcal infections
XX

PS Claim 2; Page 85-88; 160pp; English.
XX

CC The present sequence represents polypeptide R2, a new N-terminal
CC choline binding protein A (Cbpa) truncate of Streptococcus pneumoniae
CC serotype type 4. Host-vector systems for production of R2 and other
CC N-terminal Cbpa truncates are provided. Important features of R2
CC include: (1) complete correlation of bioactivity with full-length
CC Cbpa for recognition of purified glycoconjugate receptor analogues,
CC lung cells and animal models; and (2) cross-protection between type
CC 4 derived agents and bacteria in in vitro assays using other
CC serotypes (e.g. 6B and 2), important for use in vaccine, prophylaxis
CC and therapy. The invention relates generally to novel N-terminal
CC Cbpa truncates and polynucleotides encoding them, host-vector
CC systems, and antibodies that specifically bind to the truncates.
CC The invention also relates to vaccines including N-terminal Cbpa
CC truncates, which provide protection or elicit protective antibodies
CC to bacterial infection, specifically pneumococcus, and to use of
CC antibodies and adjuvants against such polypeptides in diagnosis
CC and passive immunotherapy. The polypeptides and/or polynucleotides
CC are also useful as competitive inhibitors of bacterial adhesion of
CC pneumococcus.
XX

XX Sequence 406 AA:

Query Match 78.0%; Score 1979.5; DB 20; Length 406;
Best Local Similarity 99.8%; Pred. No. 2.9e-118;
Matches 405; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 38 ENEGATQVPTSSNRANESQAEQEQPKLDSERDKARKEVEYVKYKIVGESYAKSTKKRH 97
DB 1 ENEGATQVPTSSNRANESQAEQEQPKLDSERDKARKEVEYVKYKIVGESYAKSTKKRH 60
QY 98 TITVALVNEINNINKIEYINKIVESTSESQLOILMESRSKYDEAVSKFEKSSSSSSSDS 157
DB 61 TITVALVNEINNINKIEYINKIVESTSESQLOILMESRSKYDEAVSKFEKSSSSSSSDS 120
QY 158 STKPEASDTAKPNKTEPGEKVAEAKKVEEA-KKAKDQKEEDRRNYPITYKTLELEIA 216
DB 121 STKPEASDTAKPNKTEPGEKVAEAKKVEEA-KKAKDQKEEDRRNYPITYKTLELEIA 180
QY 217 ESDVAVKKAELVYVKANEPDEQIKQAEAEVSKQAEATRLKKITDREBAEEBAKR 276
DB 181 ESDVAVKKAELVYVKANEPDEQIKQAEAEVSKQAEATRLKKITDREBAEEBAKR 240
QY 277 RADAKEQKPKGRARAGVPGELATPDKKENDAKSSDSSVGEETLPSPLKPEKKVAEAK 336
DB 241 RADAKEQKPKGRARAGVPGELATPDKKENDAKSSDSSVGEETLPSPLKPEKKVAEAK 300
QY 337 KVEAKKKKAEADQKEEDRRNYPITYKTLELEIAESDVEVKKAELELVKEAKPEPRNEBKV 396
DB 301 KVEAKKKKAEADQKEEDRRNYPITYKTLELEIAESDVEVKKAELELVKEAKPEPRNEBKV 360
QY 397 KQAKAEVSKKAATRLKKITDRKKAEEBAKRAAEEDKYKEKPA 442
DB 361 KQAKAEVSKKAATRLKKITDRKKAEEBAKRAAEEDKYKEKPA 406

Search completed: November 21, 2003, 13:31:58
Job time : 57.4631 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:32:02 ; Search time 87.0043 Seconds
(without alignments)
1882.497 Million cell updates/sec

Title: US-09-298-523C-13_COPY_263_442

Perfect score: 889
Sequence: 1 IKTDREAEAEBAKRAADAKE.....AEEBAKRAAEEDKVKKEKPA 180

Scoring table:
BLOSUM62
Gapol 10.0 , Gapext 0.5

Searched: 5728757 seqs, 909918778 residues

Total number of hits satisfying chosen parameters: 5728757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Pending Patents NA Main:*

1:	/cgn2_6/prodata/1/paa/PCRRUS_COMB.pep.*
2:	/cgn2_6/prodata/1/paa/US06_COMB.pep.*
3:	/cgn2_6/prodata/1/paa/US07_COMB.pep.*
4:	/cgn2_6/prodata/1/paa/US080_COMB.pep.*
5:	/cgn2_6/prodata/1/paa/US081_COMB.pep.*
6:	/cgn2_6/prodata/1/paa/US082_COMB.pep.*
7:	/cgn2_6/prodata/1/paa/US083_COMB.pep.*
8:	/cgn2_6/prodata/1/paa/US084_COMB.pep.*
9:	/cgn2_6/prodata/1/paa/US085_COMB.pep.*
10:	/cgn2_6/prodata/1/paa/US086_COMB.pep.*
11:	/cgn2_6/prodata/1/paa/US087_COMB.pep.*
12:	/cgn2_6/prodata/1/paa/US088_COMB.pep.*
13:	/cgn2_6/prodata/1/paa/US089_COMB.pep.*
14:	/cgn2_6/prodata/1/paa/US090_COMB.pep.*
15:	/cgn2_6/prodata/1/paa/US091_COMB.pep.*
16:	/cgn2_6/prodata/1/paa/US092_COMB.pep.*
17:	/cgn2_6/prodata/1/paa/US093_COMB.pep.*
18:	/cgn2_6/prodata/1/paa/US094_COMB.pep.*
19:	/cgn2_6/prodata/1/paa/US095_COMB.pep.*
20:	/cgn2_6/prodata/1/paa/US096_COMB.pep.*
21:	/cgn2_6/prodata/1/paa/US097A_COMB.pep.*
22:	/cgn2_6/prodata/1/paa/US097B_COMB.pep.*
23:	/cgn2_6/prodata/1/paa/US098_COMB.pep.*
24:	/cgn2_6/prodata/1/paa/US099A_COMB.pep.*
25:	/cgn2_6/prodata/1/paa/US099B_COMB.pep.*
26:	/cgn2_6/prodata/1/paa/US100_COMB.pep.*
27:	/cgn2_6/prodata/1/paa/US101_COMB.pep.*
28:	/cgn2_6/prodata/1/paa/US102_COMB.pep.*
29:	/cgn2_6/prodata/1/paa/US103_COMB.pep.*
30:	/cgn2_6/prodata/1/paa/US104_COMB.pep.*
31:	/cgn2_6/prodata/1/paa/US106_COMB.pep.*
32:	/cgn2_6/prodata/1/paa/US60_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	889	100.0	406	14	US-09-056-019-1 Sequence 1, Appl

2	889	100.0	406	14	US-09-056-019B-1	Sequence 1, Appl
3	889	100.0	406	16	US-09-287-788-1	Sequence 1, Appl
4	889	100.0	428	14	US-09-056-019-24	Sequence 24, Appl
5	889	100.0	428	14	US-09-056-019B-24	Sequence 24, Appl
6	889	100.0	428	16	US-09-287-788-24	Sequence 24, Appl
7	889	100.0	446	28	US-10-254-995-9	Sequence 9, Appl
8	889	100.0	453	13	US-08-961-083-38	Sequence 38, Appl
9	889	100.0	453	22	US-09-765-271-38	Sequence 38, Appl
10	889	100.0	453	22	US-09-765-272-38	Sequence 38, Appl
11	889	100.0	453	22	US-09-765-272A-38	Sequence 38, Appl
12	889	100.0	460	14	US-09-056-019-38	Sequence 38, Appl
13	889	100.0	460	14	US-09-056-019B-38	Sequence 38, Appl
14	889	100.0	460	16	US-09-287-788-38	Sequence 38, Appl
15	889	100.0	511	16	US-09-298-523B-13	Sequence 13, Appl
16	889	100.0	511	21	US-09-298-523C-13	Sequence 13, Appl
17	889	100.0	511	16	US-09-748-875-13	Sequence 13, Appl
18	889	100.0	511	29	US-10-341-201-13	Sequence 13, Appl
19	889	100.0	513	16	US-09-298-523B-12	Sequence 12, Appl
20	889	100.0	513	16	US-09-298-523C-12	Sequence 12, Appl
21	889	100.0	513	21	US-09-748-875-12	Sequence 12, Appl
22	889	100.0	513	29	US-10-341-201-12	Sequence 12, Appl
23	889	100.0	655	14	US-09-056-019-2	Sequence 2, Appl
24	889	100.0	655	14	US-09-056-019B-2	Sequence 2, Appl
25	889	100.0	655	16	US-09-287-788-2	Sequence 2, Appl
26	889	100.0	693	22	US-09-769-787-185	Sequence 185, Appl
27	889	100.0	693	29	US-10-333-120A-10	Sequence 10, Appl
28	884	99.4	446	28	US-10-254-995-6	Sequence 6, Appl
29	884	99.4	459	14	US-09-056-019-39	Sequence 39, Appl
30	884	99.4	459	14	US-09-056-019B-39	Sequence 39, Appl
31	884	99.4	459	16	US-09-287-788-39	Sequence 39, Appl
32	861	96.9	451	16	US-09-298-523B-67	Sequence 67, Appl
33	861	96.9	451	16	US-09-298-523C-67	Sequence 67, Appl
34	861	96.9	451	21	US-09-748-875-67	Sequence 67, Appl
35	861	96.9	451	29	US-10-341-201-67	Sequence 67, Appl
36	856	96.3	437	14	US-09-056-019B-40	Sequence 40, Appl
37	803.5	90.4	376	14	US-09-056-019-7	Sequence 7, Appl
38	803.5	90.4	376	16	US-09-287-788-7	Sequence 7, Appl
39	803.5	90.4	376	16	US-09-056-019-35	Sequence 35, Appl
40	803.5	90.4	413	14	US-09-056-019B-35	Sequence 35, Appl
41	803.5	90.4	413	16	US-09-287-788-35	Sequence 35, Appl
42	803.5	90.4	413	28	US-10-254-995-16	Sequence 16, Appl
43	803.5	90.4	581	16	US-09-298-523B-56	Sequence 56, Appl
44	803.5	90.4	581	16	US-09-298-523C-56	Sequence 56, Appl
45	803.5	90.4	581	16	US-09-298-523B-56	Sequence 56, Appl

ALIGNMENTS

RESULT 1
US-09-056-019-1 Application US/09056019A
Sequence 1, Appl
GENERAL INFORMATION:
APPLICANT: Tuomenen, Elaine I
APPLICANT: Witzmann, Theresa
APPLICANT: Maesure, H. R.
APPLICANT: Johnson, Leslie S.
TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL
TITLE OF INVENTION: CHOLESTEROL BINDING PROTEIN A TRUNCATE, VACCINE DERIVED
FILE REFERENCE: 1340-1-017
CURRENT APPLICATION NUMBER: US/09/056, 019A
CURRENT FILING DATE: 1998-04-07
NUMBER OF SEQ ID NOS: 39
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1
LENGTH: 406
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-056-019-1

Query Match 100.0%; Score 889; DB 14; Length 406;

Best Local Similarity 100.0%; Pred. No. 1.5e-49;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPDCKENDAKSSDSSVGEETLPS 60
Db IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPDCKENDAKSSDSSVGEETLPS 286
OY 61 PSLKPEKVAEAEKKVEAEAKKKAEDQKEDRRNYPNTYKTLLEIAESDVEVKKALEL 120
Db PSLKPEKVAEAEKKVEAEAKKKAEDQKEDRRNYPNTYKTLLEIAESDVEVKKALEL 346
OY 121 VKEAKEPRNEEKVQKAAEVESKKAATRLLEKIKTRDKKAAEEBAKRAAEEDVKKEKPA 180
Db VKEAKEPRNEEKVQKAAEVESKKAATRLLEKIKTRDKKAAEEBAKRAAEEDVKKEKPA 406

RESULT 2
US-09-056-019B-1
; Sequence 1, Application US/09056019B
; GENERAL INFORMATION:

APPLICANT: Tuomanen, Elaine I
APPLICANT: Witzemann, Theresa M.
APPLICANT: Masure, H. R.
APPLICANT: Johnson, Leslie S.
APPLICANT: Koenig, Scott
TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL
TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED
FILE REFERENCE: 5853-2
CURRENT APPLICATION NUMBER: US/09/056,019B
CURRENT FILING DATE: 1998-04-07
NUMBER OF SEQ ID NOS: 40
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 1
LENGTH: 406
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-056-019B-1

Query Match 100.0%; Score 889; DB 14; Length 406;
Best Local Similarity 100.0%; Pred. No. 1.5e-49;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPDCKENDAKSSDSSVGEETLPS 60
Db IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPDCKENDAKSSDSSVGEETLPS 286
OY 61 PSLKPEKVAEAEKKVEAEAKKKAEDQKEDRRNYPNTYKTLLEIAESDVEVKKALEL 120
Db PSLKPEKVAEAEKKVEAEAKKKAEDQKEDRRNYPNTYKTLLEIAESDVEVKKALEL 346
OY 121 VKEAKEPRNEEKVQKAAEVESKKAATRLLEKIKTRDKKAAEEBAKRAAEEDVKKEKPA 180
Db VKEAKEPRNEEKVQKAAEVESKKAATRLLEKIKTRDKKAAEEBAKRAAEEDVKKEKPA 406

RESULT 3
US-09-287-788-1
; Sequence 1, Application US/09287788B
; GENERAL INFORMATION:

APPLICANT: Tuomanen, Elaine I
APPLICANT: Masure, H. R.
TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL
TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED
FILE REFERENCE: 1340-1-017N
CURRENT APPLICATION NUMBER: US/09/287,788B
CURRENT FILING DATE: 1998-04-07
PRIOR APPLICATION NUMBER: 60/080,878
NUMBER OF SEQ ID NOS: 39
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 1

LENGTH: 406
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-287-788-1

Query Match 100.0%; Score 889; DB 16; Length 406;
Best Local Similarity 100.0%; Pred. No. 1.5e-49;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPDCKENDAKSSDSSVGEETLPS 60
Db IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPDCKENDAKSSDSSVGEETLPS 286
OY 61 PSLKPEKVAEAEKKVEAEAKKKAEDQKEDRRNYPNTYKTLLEIAESDVEVKKALEL 120
Db PSLKPEKVAEAEKKVEAEAKKKAEDQKEDRRNYPNTYKTLLEIAESDVEVKKALEL 346
OY 121 VKEAKEPRNEEKVQKAAEVESKKAATRLLEKIKTRDKKAAEEBAKRAAEEDVKKEKPA 180
Db VKEAKEPRNEEKVQKAAEVESKKAATRLLEKIKTRDKKAAEEBAKRAAEEDVKKEKPA 406

RESULT 4
US-09-056-019-24
; Sequence 24, Application US/09056019A
; GENERAL INFORMATION:

APPLICANT: Tuomanen, Elaine I
APPLICANT: Witzemann, Theresa
APPLICANT: Masure, H. R.
APPLICANT: Johnson, Leslie S.
APPLICANT: Koenig, Scott
TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL
TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED
FILE REFERENCE: 1340-1-017
CURRENT APPLICATION NUMBER: US/09/056,019A
CURRENT FILING DATE: 1998-04-07
NUMBER OF SEQ ID NOS: 39
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 24
LENGTH: 428
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-056-019-24

Query Match 100.0%; Score 889; DB 14; Length 428;
Best Local Similarity 100.0%; Pred. No. 1.6e-49;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPDCKENDAKSSDSSVGEETLPS 60
Db IKTDREAEAEAKRRADAKGKPKGRAGVPGELATPDCKENDAKSSDSSVGEETLPS 286
OY 61 PSLKPEKVAEAEKKVEAEAKKKAEDQKEDRRNYPNTYKTLLEIAESDVEVKKALEL 120
Db PSLKPEKVAEAEKKVEAEAKKKAEDQKEDRRNYPNTYKTLLEIAESDVEVKKALEL 346
OY 121 VKEAKEPRNEEKVQKAAEVESKKAATRLLEKIKTRDKKAAEEBAKRAAEEDVKKEKPA 180
Db VKEAKEPRNEEKVQKAAEVESKKAATRLLEKIKTRDKKAAEEBAKRAAEEDVKKEKPA 406

RESULT 5
US-09-056-019B-24
; Sequence 24, Application US/09056019B
; GENERAL INFORMATION:

APPLICANT: Tuomanen, Elaine I
APPLICANT: Witzemann, Theresa M.
APPLICANT: Masure, H. R.
APPLICANT: Johnson, Leslie S.
APPLICANT: Koenig, Scott
TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL
TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED

FILE OF INVENTION: THEREFROM AND USES THEREOF
FILE REFERENCE: 5853-2
CURRENT APPLICATION NUMBER: US/09/056,019B
CURRENT FILING DATE: 1998-04-07
NUMBER OF SEQ ID NOS: 40
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 24
LENGTH: 428
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-056-019B-24

Query Match 100.0%; Score 889; DB 14; Length 428;
Best Local Similarity 100.0%; Pred. No. 1.6e-49;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IKTDREAEAEERADAKGKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPS 60
DB 227 IKTDREAEAEERADAKGKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPS 286
QY 61 PSLEPEKVAEAEKVEAEKKAEDQKEDRRNPTNTYKTLLEIAESDVEVKAELEL 120
DB 287 PSLEPEKVAEAEKVEAEKKAEDQKEDRRNPTNTYKTLLEIAESDVEVKAELEL 346
QY 121 VKEAKEPRNEEKYKQAAEVESKKAETRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 180
DB 347 VKEAKEPRNEEKYKQAAEVESKKAETRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 406

RESULT 6
US-09-287-788-24
Sequence 24, Application US/09287788B
GENERAL INFORMATION:

APPLICANT: Tuomane, Elaine I.
TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL
TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED
FILE REFERENCE: 1340-1-017N
CURRENT APPLICATION NUMBER: US/09/287,788B
CURRENT FILING DATE: 1999-04-07
PRIOR APPLICATION NUMBER: 60/080,878
PRIOR FILING DATE: 1998-04-07
NUMBER OF SEQ ID NOS: 39
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 24
LENGTH: 428
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-287-788-24

Query Match 100.0%; Score 889; DB 16; Length 428;
Best Local Similarity 100.0%; Pred. No. 1.6e-49;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IKTDREAEAEERADAKGKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPS 60
DB 227 IKTDREAEAEERADAKGKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPS 286
QY 61 PSLEPEKVAEAEKVEAEKKAEDQKEDRRNPTNTYKTLLEIAESDVEVKAELEL 120
DB 287 PSLEPEKVAEAEKVEAEKKAEDQKEDRRNPTNTYKTLLEIAESDVEVKAELEL 346
QY 121 VKEAKEPRNEEKYKQAAEVESKKAETRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 180
DB 347 VKEAKEPRNEEKYKQAAEVESKKAETRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 406

RESULT 7
US-10-254-995-9
Sequence 9, Application US/10254995
GENERAL INFORMATION:
APPLICANT: Wizemann, Theresa M.

APPLICANT: Koenig, Scott
APPLICANT: Johnson, Leslie S
FILE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
FILE REFERENCE: 469201-396
CURRENT APPLICATION NUMBER: US/10/254,995
CURRENT FILING DATE: 2002-09-25
PRIOR APPLICATION NUMBER: US/09/286,981
PRIOR FILING DATE: 1999-04-06
PRIOR APPLICATION NUMBER: US 60/085,743
PRIOR FILING DATE: 1998-05-15
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 9
LENGTH: 446
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-10-254-995-9

Query Match 100.0%; Score 889; DB 28; Length 446;
Best Local Similarity 100.0%; Pred. No. 1.7e-49;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IKTDREAEAEERADAKGKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPS 60
DB 228 IKTDREAEAEERADAKGKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPS 287
QY 61 PSLEPEKVAEAEKVEAEKKAEDQKEDRRNPTNTYKTLLEIAESDVEVKAELEL 120
DB 288 PSLEPEKVAEAEKVEAEKKAEDQKEDRRNPTNTYKTLLEIAESDVEVKAELEL 347
QY 121 VKEAKEPRNEEKYKQAAEVESKKAETRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 180
DB 348 VKEAKEPRNEEKYKQAAEVESKKAETRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 407

RESULT 8
US-08-961-083-38
Sequence 38, Application US/08961083
GENERAL INFORMATION:

APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE: 30-Oct-1997
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: <Unknown>
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 453 amino acids
TYPE: amino acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 38:
US-09-961-083-38

Query Match 100.0%; Score 889; DB 13; Length 453;
Best Local Similarity 100.0%; Pred. No. 1.7e-49;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IKTDREAEAEAKRRADAEQGGKPKGRAGVPGELATPPDKENDAKSSDSSVGEETLPS 60
DB 25 IKTDREAEAEAEAKRRADAEQGGKPKGRAGVPGELATPPDKENDAKSSDSSVGEETLPS 84
QY 61 PSLEPEKKVAEAEKKVEAEKKKAEADQKEEDRRNPTNTYKTLLEIAESDVEVKAELEL 120
DB 85 PSLEPEKKVAEAEKKVEAEKKKAEADQKEEDRRNPTNTYKTLLEIAESDVEVKAELEL 144
QY 121 VKEAEKPRNEEKVKQAKAEVESKKAETRLKIKTDKKAEAEAKRKAABEDVKEXKA 180
DB 145 VKEAEKPRNEEKVKQAKAEVESKKAETRLKIKTDKKAEAEAKRKAABEDVKEXKA 204

RESULT 9
US-09-765-271-38

Sequence 38, Application US/09765271
GENERAL INFORMATION:

APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESSES:

ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/765,271
FILING DATE: 22-Jan-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/536,784
FILING DATE: <Unknown>
APPLICATION NUMBER: 08/961,083
FILING DATE: OCT-30-1997

ATTORNEY/AGENT INFORMATION:

NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB340P3

TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 453 amino acids
TYPE: amino acid
STRANDEDNESS: single

MOLECULE TYPE: protein
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 38:
US-09-765-271-38

Query Match 100.0%; Score 889; DB 22; Length 453;
Best Local Similarity 100.0%; Pred. No. 1.7e-49;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IKTDREAEAEAKRRADAEQGGKPKGRAGVPGELATPPDKENDAKSSDSSVGEETLPS 60

DB 25 IKTDREAEAEAKRRADAEQGGKPKGRAGVPGELATPPDKENDAKSSDSSVGEETLPS 84
QY 61 PSLEPEKKVAEAEKKVEAEKKKAEADQKEEDRRNPTNTYKTLLEIAESDVEVKAELEL 120
DB 85 PSLEPEKKVAEAEKKVEAEKKKAEADQKEEDRRNPTNTYKTLLEIAESDVEVKAELEL 144
QY 121 VKEAEKPRNEEKVKQAKAEVESKKAETRLKIKTDKKAEAEAKRKAABEDVKEXKA 180
DB 145 VKEAEKPRNEEKVKQAKAEVESKKAETRLKIKTDKKAEAEAKRKAABEDVKEXKA 204

RESULT 10
US-09-765-272-38

Sequence 38, Application US/09765272
GENERAL INFORMATION:

APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESSES:

ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/765,272
FILING DATE: 22-Jan-2001
CLASSIFICATION: <Unknown>
APPLICATION NUMBER: 08/961,083
FILING DATE: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 36,373
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 38:

SEQUENCE CHARACTERISTICS:
LENGTH: 453 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 38:
US-09-765-272-38

Query Match 100.0%; Score 889; DB 22; Length 453;
Best Local Similarity 100.0%; Pred. No. 1.7e-49;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IKTDREAEAEAKRRADAEQGGKPKGRAGVPGELATPPDKENDAKSSDSSVGEETLPS 60
DB 25 IKTDREAEAEAEAKRRADAEQGGKPKGRAGVPGELATPPDKENDAKSSDSSVGEETLPS 84
QY 61 PSLEPEKKVAEAEKKVEAEKKKAEADQKEEDRRNPTNTYKTLLEIAESDVEVKAELEL 120
DB 85 PSLEPEKKVAEAEKKVEAEKKKAEADQKEEDRRNPTNTYKTLLEIAESDVEVKAELEL 144
QY 121 VKEAEKPRNEEKVKQAKAEVESKKAETRLKIKTDKKAEAEAKRKAABEDVKEXKA 180
DB 145 VKEAEKPRNEEKVKQAKAEVESKKAETRLKIKTDKKAEAEAKRKAABEDVKEXKA 204

RESULT 11
US-09-765-272A-38

```
; Sequence 38, Application US/09765272A
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and
;                               Vaccines
; NUMBER OF SEQUENCES: 454
; CORRESPONDENCE ADDRESS:
; ADDRESS: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: Dell Latitude C610
; OPERATING SYSTEM: Windows 2000
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272A
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Lin J. Hymel
; REGISTRATION NUMBER: 45,414
; REFERENCE/DOCKET NUMBER: PB340P2C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 610-5790
; TELEFAX: (301) 309-8439
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 453 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 38:
US-09-765-272A-38

Query Match          100.0%; Score 889; DB 22; Length 453;
Best Local Similarity 100.0%; Pred. No. 1,7e-49;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IKTDREAEAEERAKRADAKGKPKGRAGVPGELATPPDKENDAKSSDSSVGEETLPS 60
DB 25 IKTDREAEAEERAKRADAKGKPKGRAGVPGELATPPDKENDAKSSDSSVGEETLPS 84
QY 61 PSLKPEKKVAEAEKKVEAKKKAADQKEEDRRNPTNTYKTLLELEIASDVEVKAELEL 120
DB 85 PSLKPEKKVAEAEKKVEAKKKAADQKEEDRRNPTNTYKTLLELEIASDVEVKAELEL 144
QY 121 VKEBAKEPRNEEKVKQAKAEVESKKAATRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 180
DB 145 VKEBAKEPRNEEKVKQAKAEVESKKAATRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 204

RESULT 12
US-09-056-019-38
; Sequence 38, Application US/09056019A
; GENERAL INFORMATION:
; APPLICANT: Tuomanen, Elaine I
; APPLICANT: Wisemann, Theresa
; APPLICANT: Masure, H. R.
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Koenig, Scott
; TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL
; TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED
; FILE REFERENCE: 1340-1-017
; CURRENT APPLICATION NUMBER: US/09/056,019A
```

```
; CURRENT FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 38
; LENGTH: 460
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-056-019-38

Query Match          100.0%; Score 889; DB 14; Length 460;
Best Local Similarity 100.0%; Pred. No. 1,7e-49;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IKTDREAEAEERAKRADAKGKPKGRAGVPGELATPPDKENDAKSSDSSVGEETLPS 60
DB 242 IKTDREAEAEERAKRADAKGKPKGRAGVPGELATPPDKENDAKSSDSSVGEETLPS 301
QY 61 PSLKPEKKVAEAEKKVEAKKKAADQKEEDRRNPTNTYKTLLELEIASDVEVKAELEL 120
DB 302 PSLKPEKKVAEAEKKVEAKKKAADQKEEDRRNPTNTYKTLLELEIASDVEVKAELEL 361
QY 121 VKEBAKEPRNEEKVKQAKAEVESKKAATRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 180
DB 362 VKEBAKEPRNEEKVKQAKAEVESKKAATRLLEKIKTRKKAEEBAKRAAEEDVKKEKPA 421

RESULT 14
US-09-287-788-38
; Sequence 38, Application US/09287788B
; GENERAL INFORMATION:
; APPLICANT: Tuomanen, Elaine I
; APPLICANT: Masure, H. R.
; TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL
```

;; TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED
;; TITLE OF INVENTION: THEREFOR AND USES THEREOF
;; FILE REFERENCE: 1340-1-017N
;; CURRENT APPLICATION NUMBER: US/09/287,788B
;; CURRENT FILING DATE: 1999-04-07
;; PRIOR APPLICATION NUMBER: 60/080,878
;; PRIOR FILING DATE: 1998-04-07
;; NUMBER OF SEQ ID NOS: 39
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 38
;; LENGTH: 460
;; TYPE: PRT
;; ORGANISM: Streptococcus pneumoniae
US-09-287-788-38

Query Match 100.0%; Score 889; DB 16; Length 460;
Best Local Similarity 100.0%; Pred. No. 1.7e-49;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IKTDREAEAEAEKRRADAEQGGKRGKRGVPGELATPDKKENDAKSSDSSVGEEETLPS 60
DB 242 IKTDREAEAEAEKRRADAEQGGKRGKRGVPGELATPDKKENDAKSSDSSVGEEETLPS 301
QY 61 PSLEPEKKVAEAEKVEAEKKAEDQKEDRRNYPNTYKTLLEIAESDVEVKKALEL 120
DB 302 PSLEPEKKVAEAEKVEAEKKAEDQKEDRRNYPNTYKTLLEIAESDVEVKKALEL 361
QY 121 VKEEAKPRNEEKVQAQAEVSKAEATRLKIKTDRKAAEEAKRKAEEEDKYEKPA 180
DB 362 VKEEAKPRNEEKVQAQAEVSKAEATRLKIKTDRKAAEEAKRKAEEEDKYEKPA 421

RESULT 15
US-09-298-523B-13
; Sequence 13, Application US/09298523B
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/298,523B
; CURRENT FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-298-523B-13

Query Match 100.0%; Score 889; DB 16; Length 511;
Best Local Similarity 100.0%; Pred. No. 2e-49;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IKTDREAEAEAEKRRADAEQGGKRGKRGVPGELATPDKKENDAKSSDSSVGEEETLPS 60
DB 263 IKTDREAEAEAEKRRADAEQGGKRGKRGVPGELATPDKKENDAKSSDSSVGEEETLPS 322
QY 61 PSLEPEKKVAEAEKVEAEKKAEDQKEDRRNYPNTYKTLLEIAESDVEVKKALEL 120
DB 323 PSLEPEKKVAEAEKVEAEKKAEDQKEDRRNYPNTYKTLLEIAESDVEVKKALEL 382
QY 121 VKEEAKPRNEEKVQAQAEVSKAEATRLKIKTDRKAAEEAKRKAEEEDKYEKPA 180
DB 383 VKEEAKPRNEEKVQAQAEVSKAEATRLKIKTDRKAAEEAKRKAEEEDKYEKPA 442

Search completed: November 21, 2003, 13:40:58
Job time : 88.0043 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:32:37 ; Search time 7.03329 Seconds
(without alignments)
1285.946 Million cell updates/sec

Title: US-09-298-523C-13_COPY_263_442
Perfect score: 889
Sequence: 1 IKTDREAEAEKRRADAKE.....AEEKAKKAEEDKVEKXA 180

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 261868-seqs, 50246823 residues

Total number of hits satisfying chosen parameters: 261868

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Pending_Patents_AA_New:*
1: /cgn2_6/prodata/1/paa/PCT_NEW_COMB.pep:*
2: /cgn2_6/prodata/1/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/prodata/1/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/prodata/1/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/prodata/1/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/prodata/1/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/prodata/1/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	889	100.0	693	1	PCT-US03-27401-468 Sequence 468, App
2	889	100.0	693	6	US-10-472-928-4598 Sequence 4598, App
3	187.5	21.1	1026	6	US-10-258-899A-1487 Sequence 1487, App
4	187.5	21.1	1033	6	US-10-258-899A-3455 Sequence 3455, App
5	162.5	18.3	507	6	US-10-425-114A-47285 Sequence 47285, App
6	158.5	17.8	257	1	PCT-US03-11802-20 Sequence 20, App
7	158.5	17.8	487	1	PCT-US03-11802-21 Sequence 21, App
8	156	17.5	744	1	PCT-US03-27401-251 Sequence 251, App
9	156	17.5	744	6	US-10-472-928-32 Sequence 32, App
10	154.5	17.4	1533	6	US-10-679-063-23311 Sequence 23311, App
11	152	17.1	793	7	US-60-490-890-2605 Sequence 2605, App
12	148	16.6	725	6	US-10-258-899A-1881 Sequence 1881, App
13	148	16.6	774	6	US-10-258-899A-3849 Sequence 3849, App
14	146	16.4	369	5	US-09-820-843B-95 Sequence 95, App
15	146	16.4	665	5	US-09-820-843B-107 Sequence 107, App
16	146	16.4	938	7	US-60-490-890-495 Sequence 495, App
17	145.5	16.4	356	5	US-09-820-843B-27 Sequence 27, App
18	143.5	16.1	4684	1	PCT-US03-14382-432 Sequence 432, App
19	143.5	16.1	4684	1	PCT-US03-14382-434 Sequence 434, App
20	143.5	16.1	4684	1	PCT-US03-14382-432 Sequence 432, App
21	143.5	16.1	4684	1	PCT-US03-14382A-434 Sequence 434, App
22	143.5	16.1	4684	7	US-60-479-073-497 Sequence 497, App
23	143.5	16.1	4684	7	US-60-479-073-506 Sequence 506, App
24	141	15.9	612	7	US-60-500-337-2473 Sequence 2473, App
25	141	15.9	612	7	US-60-500-337-2475 Sequence 2475, App
26	141	15.9	662	7	US-60-500-337-2476 Sequence 2476, App

27	140	15.7	2524	7	US-60-502-656-219 Sequence 219, App
28	140	15.7	2524	7	US-60-512-690-253 Sequence 253, App
29	140	15.7	4574	7	US-60-479-073-495 Sequence 495, App
30	140	15.7	4576	1	PCT-US03-3147C-281 Sequence 281, App
31	139.5	15.7	916	7	US-60-490-890-475 Sequence 475, App
32	138.5	15.6	372	5	US-09-820-843B-8 Sequence 8, App
33	138	15.5	204	1	PCT-US03-26864-304 Sequence 304, App
34	138	15.5	204	1	PCT-US03-27107-304 Sequence 304, App
35	138	15.5	572	6	US-10-425-114A-43212 Sequence 43212, App
36	137.5	15.5	486	6	US-10-679-063-14456 Sequence 14456, App
37	137.5	15.5	543	6	US-10-679-063-14127 Sequence 14127, App
38	137.5	15.5	926	5	US-09-897-516A-6173 Sequence 6173, App
39	137	15.4	558	7	US-60-490-890-2607 Sequence 2607, App
40	137	15.4	564	7	US-60-490-890-2611 Sequence 2611, App
41	137	15.4	1027	7	US-60-495-114-1376 Sequence 1376, App
42	136.5	15.4	284	7	US-60-485-450-1572 Sequence 1572, App
43	136.5	15.4	515	6	US-10-343-650A-418 Sequence 418, App
44	136.5	15.4	1307	6	US-10-258-899A-3084 Sequence 3084, App
45	136.5	15.4	1307	6	US-10-258-899A-3085 Sequence 3085, App

ALIGNMENTS

RESULT 1
PCT-US03-27401-468
Sequence 468, Application PC/TUS0327401
GENERAL INFORMATION:
APPLICANT: TUFTS UNIVERSITY
TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSIS, TREATMENT AND
FILE REFERENCE: 700355-52941-PCT
CURRENT APPLICATION NUMBER: PCT/US03/27401
CURRENT FILING DATE: 2003-09-02
PRIOR FILING DATE: 2002-08-30
NUMBER OF SEQ ID NOS: 560
SOFTWARE: PatentIn version 3.2
SEQ ID NO 468
LENGTH: 693
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
PCT-US03-27401-468

Query Match 100.0%; Score 889; DB 1; Length 693;
Best Local Similarity 100.0%; Pred. No. 5.1e-32;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IKTDREAEAEKRRADAKEQKPKGRAGVPGELATPDKKENDAKSSDSVGEETLPS 60
DB 265 IKTDREAEAEKRRADAKEQKPKGRAGVPGELATPDKKENDAKSSDSVGEETLPS 324

QY 61 PSLEKEKVAEAEKVEAEKKAEDQKEDRRNYPNTYKTELEIAESDVEYKAELEI 120
DB 325 PSLEKEKVAEAEKVEAEKKAEDQKEDRRNYPNTYKTELEIAESDVEYKAELEI 384

QY 121 VEEAEKPEPNEKVKQAKAEVSSKAEATRELEKIKTDKKAEEAEKKAEEKVEKXA 180
DB 385 VEEAEKPEPNEKVKQAKAEVSSKAEATRELEKIKTDKKAEEAEKKAEEKVEKXA 444

RESULT 2
US-10-472-928-4598
Sequence 4598, Application US/10472928
GENERAL INFORMATION:
APPLICANT: CHIRON SPA
TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
FILE REFERENCE: P026926WO
CURRENT APPLICATION NUMBER: US/10-472, 928
CURRENT FILING DATE: 2003-09-26
PRIOR APPLICATION NUMBER: GB-0107658.7
PRIOR FILING DATE: 2001-03-27

```
NUMBER OF SEQ ID NOS: 4979
SOFTWARE: SeqWin99, version 1.03
SEQ ID NO 4598
LENGTH: 693
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
FEATURE:
OTHER INFORMATION: choline binding protein A (cbpa)
OTHER INFORMATION: Cellular location: outside
OTHER INFORMATION: Feature of note: WY motif
OTHER INFORMATION: Similar to strain R6 sequence 15904036 (0.E+01)
US-10-472-928-4598
```

```
Query Match          100.0%; Score 889; DB 6; Length 693;
Best Local Similarity 100.0%; Pred. No. 5.1e-32;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 IKTREAEAEAKRADAKEDQKPKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPS 60
DB 265 IKTREAEAEAKRADAKEDQKPKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPS 324
QY 61 PSLKPEKVAEAEKKEBEAKKADQKEDRRNYPNTYKTLLEIAESDVEVKKAIEEL 120
DB 325 PSLKPEKVAEAEKKEBEAKKADQKEDRRNYPNTYKTLLEIAESDVEVKKAIEEL 384
QY 121 VKEAKEPRNEKVKQAKAEVESKKAATRLKIKTDKKAEEAKKRAAEEDVKKEKPA 180
DB 385 VKEAKEPRNEKVKQAKAEVESKKAATRLKIKTDKKAEEAKKRAAEEDVKKEKPA 444
```

RESULT 3

US-10-258-899A-1487

Sequence 1487, Application US/10258899A

GENERAL INFORMATION:

```
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Drianae, Radoje T.
APPLICANT: Asundi, Vinod
APPLICANT: Zhou, Ping
APPLICANT: Xu, Chongjun
APPLICANT: Cao, Yicheng
APPLICANT: Ma, Yungqing
APPLICANT: Zhao, Qing A.
APPLICANT: Wang, Duntui
APPLICANT: Wang, Jian-Rui
APPLICANT: Ren, Feiyan
APPLICANT: Chen, Rui-hong
APPLICANT: Wang, Zhi Wei
APPLICANT: Xue, Aiding
APPLICANT: Yang, Yonghong
APPLICANT: Wehman, Tom
APPLICANT: Goodrich, Ryle
TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
FILE REFERENCE: 787CIP2-2G/US
CURRENT APPLICATION NUMBER: US/10/258,899A
CURRENT FILING DATE: 2003-11-06
PRIOR APPLICATION NUMBER: PCT/US01/04098
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: 09/774,434
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 09/728,422
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: 09/693,325
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 09/663,561
PRIOR FILING DATE: 2000-09-15
PRIOR APPLICATION NUMBER: 09/654,936
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 09/620,325
PRIOR FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 09/598,075
PRIOR FILING DATE: 2000-06-20
```

```
PRIOR APPLICATION NUMBER: 09/560,875
PRIOR FILING DATE: 2000-04-27
PRIOR APPLICATION NUMBER: 08/496,914
PRIOR FILING DATE: 2000-02-03
NUMBER OF SEQ ID NOS: 3960
SOFTWARE: Custom
SEQ ID NO 1487
LENGTH: 1026
TYPE: PRT
ORGANISM: Homo sapiens
US-10-258-899A-1487
```

```
Query Match          21.1%; Score 187.5; DB 6; Length 1026;
Best Local Similarity 31.9%; Pred. No. 0.075;
Matches 60; Conservative 25; Mismatches 76; Indels 23; Gaps 7;
```

```
QY 5 REAEAEAKRADAKEDQKPKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPS 64
DB 705 KEAEKPEKAKSPVKEAKSPKEAKSPVKEAKTPEKAKSPVKEAKS-----PEKAKS 758
QY 65 PEKVVAEAEKKEBEAKKADQKEDRRNYPNTYKTLLEI-----AESDV--EVKAE 117
DB 759 PE-KAKTLDVKSPKAPKPAKEARSPADKPEKAKSPVKEEVKSPKAKSPLEKADAKPE 817
QY 118 LEL-VKEAKEPRNE-----KVQAKAEVESKKAATRLKIKTDKKAEEAKKRAA 170
DB 818 KEIKKEEVKSPVKEEKPOEVKVPKAEKAPATKTEBKDSK--EEAPKKA 875
QY 171 EEDVKKEK 178
DB 876 PKPVKEK 883
```

RESULT 4

US-10-258-899A-3455

Sequence 3455, Application US/10258899A

GENERAL INFORMATION:

```
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Drianae, Radoje T.
APPLICANT: Asundi, Vinod
APPLICANT: Zhou, Ping
APPLICANT: Xu, Chongjun
APPLICANT: Cao, Yicheng
APPLICANT: Ma, Yungqing
APPLICANT: Zhao, Qing A.
APPLICANT: Wang, Duntui
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Chen, Rui-hong
APPLICANT: Wang, Zhi Wei
APPLICANT: Xue, Aiding
APPLICANT: Yang, Yonghong
APPLICANT: Wehman, Tom
APPLICANT: Goodrich, Ryle
TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
FILE REFERENCE: 787CIP2-2G/US
CURRENT APPLICATION NUMBER: US/10/258,899A
CURRENT FILING DATE: 2003-11-06
PRIOR APPLICATION NUMBER: PCT/US01/04098
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: 09/774,434
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 09/728,422
PRIOR FILING DATE: 2000-11-30
PRIOR APPLICATION NUMBER: 09/693,325
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 09/663,561
PRIOR FILING DATE: 2000-09-15
PRIOR APPLICATION NUMBER: 09/654,936
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 09/620,325
```

;; PRIOR FILING DATE: 2000-07-19
;; PRIOR APPLICATION NUMBER: 09/598,075
;; PRIOR FILING DATE: 2000-06-20
;; PRIOR APPLICATION NUMBER: 09/560,875
;; PRIOR FILING DATE: 2000-04-27
;; PRIOR APPLICATION NUMBER: 09/496,914
;; PRIOR FILING DATE: 2000-02-03
;; NUMBER OF SEQ ID NOS: 3960
;; SOFTWARE: Custom
;; SEQ ID NO: 3455
;; LENGTH: 1033
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-10-258-899A-3455

Query Match 21.1%; Score 187.5; DB 6; Length 1033;
Best Local Similarity 31.9%; Pred. No. 0.075;

Matches 60; Conservative 29; Mismatches 76; Indels 23; Gaps 7;

QY 5 REEAEEAKRRADAKEOGKPKGRAGVPGELATPPDKKENDAKSSDSVGEETLPSPLK 64
DB 712 KEBAKSPKAKSPVKEBAKSPKAKSPVKEAKTPPEKAKSPVKEBAKSPKAKS-----PEKAKS 765
QY 65 PEKVAEAKKVEBAKKAEDQKEEDRNYPNTYKLELEI-----AESDV--EVKAE 117
DB 766 PE-KAKTIDVKSPEAKTPAKEEARSPADKPEKAKSPVKEEVKSPKAKSPLKEDAKAPE 824
QY 118 LEL-VKEAKPRNE-----KYQAKAEVSKKAEATRLKIKTDKKAEEBAKKA 170
DB 825 KEIPKKEEVKSPVKEEKPQEVKVEPKKAEKKAEPATPTBEKKSKK--EEAPKKEA 882
QY 171 EEDVKKEK 178
DB 883 PKPKVEEK 890

RESULT 5
US-10-425-114A-47285

Sequence 47285, Application US/10425114A
GENERAL INFORMATION:

APPLICANT: Liu, Jingdong
APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E
APPLICANT: Tabaska, Jack E
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53113)B
CURRENT APPLICATION NUMBER: US/10/425,114A
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 47285
LENGTH: 507
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: 700863165_F11.pep
US-10-425-114A-47285

Query Match 18.3%; Score 162.5; DB 6; Length 507;
Best Local Similarity 28.2%; Pred. No. 0.56;

Matches 55; Conservative 39; Mismatches 76; Indels 25; Gaps 6;

QY 2 KTDREAEBAKRRADAKEOGKPKGRAGVPGELATPPDKKENDAKSSDSVGE-----56
DB 262 KEERKEEBAKRRQOEKREKRGEEBAKREBEBAKME--KERKGEBAEREEBAARR 320
QY 57 -----TLPSLSLKEKKVAEAKKVEBAKKAEDQKEEDRNYPNTYKLELEIASED 110
DB 321 EOEWRKKEBAARBOEKERRRQOEBAKKAEEKAEKEDQ-----ATREAEKQRE 374
QY 111 VEVKKAEL-----ELVKEBAKPRNEEKVKQAKAEVE--SKKAEATRLKIKTDKKA 163

DB 375 OEAKREAAQREGBAKREGRRRREEVAGKGEEEEHAGHBRKKEEFTARBEQOE 434
QY 164 EAKKAAEEDVKKEK 178
DB 435 QAERQOQKREKRROK 449

RESULT 6

PCT-US03-11802-20
Sequence 20, Application PC/TUS0311802
GENERAL INFORMATION:

APPLICANT: Washington University
TITLE OF INVENTION: Regulated Attenuation of Live Vaccines to Enhance Cross-Protectiv
FILE REFERENCE: 56029-40434
CURRENT APPLICATION NUMBER: PCT/US03/11802
CURRENT FILING DATE: 2003-04-15
PRIOR APPLICATION NUMBER: US 60/373,626
PRIOR FILING DATE: 2002-04-18
PRIOR APPLICATION NUMBER: US 60/372,616
NUMBER OF SEQ ID NOS: 21
SOFTWARE: PatentIn version 3.1
SEQ ID NO 20
LENGTH: 257
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
PCT-US03-11802-20

Query Match 17.8%; Score 158.5; DB 1; Length 257;
Best Local Similarity 29.5%; Pred. No. 0.55;

Matches 56; Conservative 43; Mismatches 50; Indels 41; Gaps 10;

QY 2 KTDREAE-----EBAKRRADAKEOGKPKGRAGV-----PELATPPDKKENDAKSSDSV 54
DB 87 KAAKDADKMIIDEKKR-----EEBAKTKFNTVRAMVPEPEQLAETKKKSEBAKQK-----138
QY 55 EETLPSLSLKEKKVAEAKKVEBAKKAEDQKEE--DRRNYPNTYKLELEIASEDEV 113
DB 139 -----APEL--TKLEBAKAKLEBAEKATBAKQVDAEEVAP-----QAKIAELENOV 185
QY 114 KKALELEVKEBAKPRNEEKVKQK-----KAIVESKKAETRLKIKTDKKAEE 163
DB 186 HRLBOEL--KEIDSESEDIYAKGFRAPLOS KDAKAKLSKLELSDKIDELDAETAKL 243
QY 164 EAKKAAEED 173
DB 244 EDLKAKEEN 253

RESULT 7

PCT-US03-11802-21
Sequence 21, Application PC/TUS0311802
GENERAL INFORMATION:

APPLICANT: Washington University
TITLE OF INVENTION: Regulated Attenuation of Live Vaccines to Enhance Cross-Protectiv
FILE REFERENCE: 56029-40434
CURRENT APPLICATION NUMBER: PCT/US03/11802
CURRENT FILING DATE: 2003-04-15
PRIOR APPLICATION NUMBER: US 60/373,626
PRIOR FILING DATE: 2002-04-18
PRIOR APPLICATION NUMBER: US 60/372,616
NUMBER OF SEQ ID NOS: 21
SOFTWARE: PatentIn version 3.1
SEQ ID NO 21
LENGTH: 467
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
PCT-US03-11802-21

	Query Match	17.8%	Score 158.5:	DB 1;	Length 487;
	Best Local Similarity	29.5%;	Pred. No. 0.82;		
	Matches	56;	Conservative	43;	Mismatches 50; Indels 41; Gaps 10.
Oy	2	KTDREAE--EAKRRADAKEOGPKGRAGKGV--PGLATPDKKENDAKSSDSVG	54		
		: : : : : :	:	:	:
Db	317	KAQDAADKMIDEAKKR--EEAATKTFTYAVAMVPEEGQAETRKKESEAKQR----	368		
Oy	55	EETLPSPSLKPPKKVAEAKKYEEKKKAEDQKE--DRRNYYNTNFKYLEIASDVEY	113		
		: : : : : :	:	:	:
Db	369	--APEL--TKLIEEAKKLEEKAKATEKKQCVDAEEVAP-----QAATALEENOV	415		
Oy	114	KKAELELVKEAEKEPRNEEKVKQA-----KAVESEKKAETATLEKI-----KTRDKKAAE	163		
		: : : : : :	:	:	:
Db	416	HRLBEL--KEIDSESEDYAKGEGRAPLQSULDAKKAKLSLTELSDKIDBLDLAELIKL	473		
Oy	164	EAKRKAABD	173		
		: : : : :	:	:	:
Db	474	EDOLCAEEN	483		

```

RESULT 8
PCT-US03-27401-251
Sequence 251, Application PC/TUS0327401
GENERAL INFORMATION:
APPLICANT: TUFTS UNIVERSITY
TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSIS, TREATMENT AND
TITLE OF INVENTION: PREVENTION OF ACTIVE INFECTION
FILE REFERENCE: 700355-52941-PCt
CURRENT APPLICATION NUMBER: PCT/US03/27401
CURRENT FILING DATE: 2003-09-02
PRIOR APPLICATION NUMBER: US 60/407,082
PRIOR FILING DATE: 2002-08-30
NUMBER OF SEQ ID NOS: 560
SOFTWARE: PatentIn version 3.2
SEQ ID NO 251
LENGTH: 744
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
PCT-US03-27401-251

```

[illegible]

```

RESULT 9
US-10-472-928-32
; Sequence 32, Application US/10472928
;
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS

```

```

, FILE REFERENCE: P0269326MO
, CURRENT APPLICATION NUMBER: US/10/472,928
, CURRENT FILING DATE: 2003-09-26
, PRIOR APPLICATION NUMBER: GB-0107658.7
, PRIOR FILING DATE: 2001-03-27
, NUMBER OF SEQ ID NOS: 4979
, SOFTWARE: SeqMan99, version 1.03
, SEQ ID NO 32
, LENGTH: 744
, TYPE: PRT
, ORGANISM: Streptococcus pneumoniae
, FEATURE:
, OTHER INFORMATION: pneumococcal surface protein A (pspa)
, OTHER INFORMATION: Cellular location: outside
, OTHER INFORMATION: Feature of note: WY motif
, OTHER INFORMATION: Similar to strain R6 sequence 15902165 (e-179)
US-10-472-928-32

```

Query Match	17.5%	Score 156;	DB 6;	Length 744;
Best Local Similarity	21.8%	Pred. No. 1.4;		
Matches	60;	Conservative	46;	Mismatches 67; Indels 102; Gaps 9
QY	2	KTDREAEAEAKRRADAEQCKPKGRAGVGPGLATPDPKKENDAKSSDSSVGEETLSP	61	
Db	40	KSLSEKKEEAKAKADTAKDYETAKKAEAEQCKYEEDQKRTKEAKAEKAEASQKLVNDV	99	
QY	62	SL-----KPEKTVAAEAKVKE-----	77	
Db	100	ALVYQNAKYEREYQNOCKYKSDAEYQKLTLEVDSKIEKARKKEQODLQNKFNVEIRAVVV	159	
QY	78	-----EAKKAEQCKED---RRNYFTNTYKLTLELAEISDVEYVKAQELLEVVE---	124	
Db	160	PEPNALETAKKAEAEAKAEKVEAKRY---DYATLKVALAKEVAKLEIEIKQYETIST	216	
QY	125	-----AKEPNN-----EEKVQKAQAEVSKAAE-----ATRLERI-	154	
Db	217	LEGEVATQHQVDNINKLLAGADPDQGEVIEIAKLKKEAEELNMQAQLAKQGTLEKTL	276	
QY	155	-----KTDKKAEAEAKRKAAB-EDKVKE	177	
Db	277	DSLDPGKQTQDELDKAEAEALDKKADLQNKVAD	311	

```

RESULT 10
US-10-679-063-23311
; Sequence 23311, Application US/10679063
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
; FILE REFERENCE: 39-15(52054)B
; CURRENT APPLICATION NUMBER: US/10/679,063
; CURRENT FILING DATE: 2003-10-02
; PRIOR APPLICATION NUMBER: 60/415,758
; PRIOR FILING DATE: 2002-10-02
; NUMBER OF SEQ ID NOS: 27373
; SEQ ID NO 23311
; LENGTH: 1533
; TYPE: PRT
; ORGANISM: Neurospora crassa
US-10-679-063-23311

```

	Query Match	17.4%	Score 154.5;	DB 6;	Length 1533;	
	Best Local Similarity	28.8%;	Pred. No. 2.5;			
	Matches	55;	Conservative	33;	Mismatches	72;
					Indels	31;
					Gaps	7
<hr/>						
QY	2 KTDREAEAEKRR---	ADAKGCKPPEGR--	AKGVGELATIPPKENDAKSSSSVGE	55		
	:	: : : :	:	:	:	:
	:	:	:	:	:	:
	:	:	:	:	:	:
	:	:	:	:	:	:
Db	1031 RAERERARAERKRQAEEEDARREAREFOAKLAAEECGAPAVASP-----	KADS---R	1074			
<hr/>						
QY	56 ETLSPSLAKEPKVAEAKEA	KKKAADQKEENRYNPNTTYTLE-	-L	104		
Db	1080 PPPPPPSKACK--ADDRRDDEAAKKAEEERLLEEREEOERTETELESASAKAOEDELAK	1133				

```

CURRENT FILING DATE: 2003-11-06
PRIORITY APPLICATION NUMBER: PCT/US01/04098
PRIORITY FILING DATE: 2001-02-05
PRIORITY APPLICATION NUMBER: 09/774,434
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: 09/772,422
PRIORITY FILING DATE: 2000-11-30
PRIORITY APPLICATION NUMBER: 09/693,325
PRIORITY FILING DATE: 2000-10-20
PRIORITY APPLICATION NUMBER: 09/663,561
PRIORITY FILING DATE: 2000-09-15
PRIORITY APPLICATION NUMBER: 09/654,936
PRIORITY FILING DATE: 2000-09-01
PRIORITY APPLICATION NUMBER: 09/620,325
PRIORITY FILING DATE: 2000-07-19
PRIORITY APPLICATION NUMBER: 09/598,075
PRIORITY FILING DATE: 2000-06-20
PRIORITY APPLICATION NUMBER: 09/566,875
PRIORITY FILING DATE: 2000-04-27
PRIORITY APPLICATION NUMBER: 09/496,914
PRIORITY FILING DATE: 2000-02-03
NUMBER OF SEQ ID NOS: 3960
SOFTWARE: Custom
SEQ ID NO 1881
LENGTH: 725
TYPE: PRT
ORGANISM: Homo sapiens
US-10-258-899A-1881

Query Match      16.6%; Score 148; DB 6; Length 725;
Best Local Similarity 25.5%; Pred. No. 3;
Matches 52; Conservative 47; Mismatches 63; Indels 42; Gaps 8

QY      1  IKTRDEAEAEAKRRADAKGCGKPKGAKGVPGLATPD---KKENDAKSSDSVVG- 54
Db      61  LNDIFEAANDSDSQOSTKEDTSSKKKKKKLROEESPDLLKKKKAKAGLTKDKSKDL 120
QY      55  EETLPSP--SLPEPKVAEAEKKEVEAKKKAEDQKEEDRRNYPNTYKTLLEIAESDVE 112
Db      121  ESSLESIVPDLRTKGRISEAKEBELKSKPKKDEVKE-----TKELK-----K 163
QY      113  VKRAELELVK---EAKEPNNEKVAQAKAEVSSKAA-----EATRLKIKTRD 158
Db      164  VKKGEIIDLTKTRREDPKENRKTKEKEFVESQVSESSVLNDSPPEDDSGLHSRRE 223
QY      159  KKAEEAEKRAAEEDKYK---EKP 179
Db      224  KONTKSARERRAQDMGLEHGEKP 247

RESULT 13
US-10-258-899A-3849
Sequence 3849, Application US/10258899A
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenghua
APPLICANT: Dirmnac, Radoje T.
APPLICANT: Asundi, Vinod
APPLICANT: Zhou, Ping
APPLICANT: Xu, Chongjun
APPLICANT: Cao, Yicheng
APPLICANT: Ma, Yungqing
APPLICANT: Zhao, Qing A.
APPLICANT: Wang, Dunrui
APPLICANT: Wang, Jian-Rui
APPLICANT: Zhang, Jie
APPLICANT: Ren, Feiyan
APPLICANT: Chen, Rui-hong
APPLICANT: Wang, Zhi Wei
APPLICANT: Xue, Aiding
APPLICANT: Yang, Yonghong
APPLICANT: Wehrman, Tom
APPLICANT: Goodrich, Ryle

```

```

; TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
; FILE REFERENCE: 787CIP2-26/US
; CURRENT APPLICATION NUMBER: US/10/258,899A
; CURRENT FILING DATE: 2003-11-06
; PRIOR APPLICATION NUMBER: PCT/US01/04098
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 09/774,434
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 09/728,422
; PRIOR FILING DATE: 2000-11-30
; PRIOR APPLICATION NUMBER: 09/693,325
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 09/663,561
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: 09/654,936
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: 09/620,325
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/598,075
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 09/496,914
; PRIOR FILING DATE: 2000-02-03
; NUMBER OF SEQ ID NOS: 3960
; SOFTWARE: Custom
; SEQ ID NO: 3849
; LENGTH: 774
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(774)
; OTHER INFORMATION: Kaa = any amino acid or nothing
US-10-258-899A-3849
```

```

Query Match      16.6%; Score 148; DB 6; Length 774;
Best Local Similarity 25.5%; Pred. No. 3.1;
Matches 52; Conservative 47; Mismatches 63; Indels 42; Gaps 8;
```

```

Qy 1 KTDREAEAEAKRRADAEQGPGRAGVPGELATPD---KKENDAKSDSSVG-- 54
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 152 LNDIDFEANSDDQSEFTKEDTSPKPKKKKLRQREKSPDDLKKAAGAKLKDKSKPDL 211
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 55 EETLPSD--SLKPEKKVAEAEKKVSEAKKKAEDQKEEDRRNYPNTYTLLELAESVVE 112
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 212 ESSLESFLVDLRTKRRISEAEKELSKPKKDEYK-----TKELK-----K 254
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 113 VKKAELVLV---EAEKPRNEEKVKQAKAEVSKKA-----EATRLKIKITDR 158
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 255 VKKGEIRDLKTRDPKRNKTKKEKVESESVLNDSPPEDDSGHSDSREE 314
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 159 KKAEEAKRKAEDDKV--EKP 179
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 315 KONTKARERAGQDWGLEHGEKFP 338
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

```

RESULT 14
US-09-820-843B-95
; Sequence 95, Application US/09820843B
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE
; FILE REFERENCE: 063915
; CURRENT APPLICATION NUMBER: US/09/820,843B
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 95
; LENGTH: 369
; TYPE: PRT
; ORGANISM: T. pallidum
```

```

; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: conserved hypothetical protein
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: g1|3322751
US-09-820-843B-95
```

```

Query Match      16.4%; Score 146; DB 5; Length 369;
Best Local Similarity 29.7%; Pred. No. 2.4;
Matches 52; Conservative 29; Mismatches 56; Indels 38; Gaps 7;
```

```

Qy 5 REAEAEAKRRADAEQGPGRAGVPGELATPDKKENDAKSDSSVGSETLPSPLK 64
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 193 RKEAE--ARRKEAEARRKAEAEARRKE--EARRKAEARRKAEARRKAE 246
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 65 PEKKVAE--AEKVEAEAKKAEDQKEEDRRNYPNTYTLLELAESDVEYKAELELVK 122
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 247 ARKAEAEARRKAEAEARRK--EAEARR-----KEAE-EARR 281
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 123 EAEKPRNEEKVKQAKAEVSKKAATRLKIKITDRKAEAEAKRKAEDDKVE 177
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 282 KEAEAEARRK--EAEARRKAEAEARRKAEAEARRKAEAEARRKAE 331
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

RESULT 15

US-09-820-843B-107

```

; Sequence 107, Application US/09820843B
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE
; FILE REFERENCE: 063915
; CURRENT APPLICATION NUMBER: US/09/820,843B
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 107
; LENGTH: 665
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: hypothetical protein
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: g1|3845248
US-09-820-843B-107
```

```

Query Match      16.4%; Score 146; DB 5; Length 665;
Best Local Similarity 24.4%; Pred. No. 3.4;
Matches 52; Conservative 39; Mismatches 82; Indels 40; Gaps 6;
```

```

Qy 2 KTDREAEAEAKRRADAEQGPGRAGVPGELATPDKKENDAKSDSSVGSETLPSP 61
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 159 KKGQDINSNAENKQDYKGVKLEBEKK--BEKISDQKVEENKKSDDHKVEENKKS 215
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 62 SLKPEKKVAEAEKKVEAEAK--KAEDQKEEDRR-----NYTNTYKTLLE--E 105
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 216 DHKVEENKKSDDHKIEEVKVEEHEDEEDKKEKSENKKNQDENEDNDEISDED 275
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 106 IAEVDV-----VKKAELVLVEAEKPEPRNEEKVKQAKAEVSKK 145
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 276 EVDVDVEDKRNENDDIDDKKETDKTLLEENENIIKEFSDDKQGNKNDTKKE--KSKD 334
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 146 AEATRLKIKITDRKAEAEAKRKAEDDKVEK 178
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 335 TEKEKSKDIEKSKDKKEKSKDKKEKKGDK 367
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

Search completed: November 21, 2003, 13:41:39
Job time : 8.03329 secs

KM	Intermediate filament; Coiled coil; Neutrone; Phosphorylation.
FT	DOMAIN 1 100 HEAD. 101 413 ROD. 101 414 TAIL. 101 1020 TAIL. 101 132 COIL 1A. 133 145 LINKER 1. 133 145 COIL 1B. 146 244 COIL 1B. 245 266 LINKER 12. 267 288 COIL 2A. 289 292 LINKER 2. 293 413 COIL 2B.
SO	SEQUENCE 1020 AA; 111780 MW; 117769DCB3DCFID4 CRC64;
Query Match	21.2%; Score 188.5; DB 1; Length 1020;
Best Local Similarity	30.9%; Pred. No. 0.011;
Matches	58; Conservative 28; Mismatches 79; Indels 23; Gaps 6
QY	5 REAAEEARRRADAEQGCPKGRAKGVGEGELATDPKKENDAKSSSVGEETLPSPSLK 64
DB	699 KEAAASPEPAKSPVKEEAKSPPEKAAPSVEEAATTPPKAKSPVKEEAKS-----PEKAKS 752
QY	65 PEKTAABAEKVBEBAKKADOKEDRRNVPTTNTYLELEIAESD-----VEVKAAE 117
DB	753 PE-KAKTLDVKSPEEATTPAKEEKRSADYFPEKAKSPVEEVKSPPEKAKSPLEKADKAP 811
QY	118 LEL-VKEAKEPRNE-----KVQAKAVESSKKAATRLKIKTKRKAEBEARKAA 170
DB	812 KEIPKKEBKSPVKEBEKPQEVBKVPPEPKAEBEKAPAIPTEEEKDSKK--EAVPKKEA 869
QY	171 EEDKYKEK 178
DB	870 PKPVKEEK 877
RESULT 2	
CALD_CHICK	STANDARD; PRT; 771 AA.
ID_CALD_CHICK	P12957; Q03698; Q90756; Q90761; Q92018; Q99230;
AC	01-JAN-1990 (Rel. 13, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DT	28-FEB-2003 (Rel. 41, Last annotation update)
DE	Caldesmon (CDM).
GN	CALDI OR CAD.
OS	Gallus gallus (Chicken).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Archosauria; Aves; Neognathae; Galliformes; Phaethonidae; Phaethoniace;
OC	Gallus.
OX	NCBI_TaxID=9031;
RN	(1)
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORM GIZZARD H-CAD).
RC	TISSUE=gizzard; And Oviduct;
RX	MEDLINE=893140480; PubMed=2760048;
RA	Bryan J., Imai M., Lee R., Moore P., Cook R.G., Lin W.-G.;
RT	"Cloning and expression of a smooth muscle caldesmon.";
RL	J. Biol. Chem. 264:13873-13879(1989).
RN	(2)
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORM GIZZARD H-CAD).
RC	TISSUE=gizzard;
RX	MEDLINE=90026426; PubMed=2803315;
RA	Hayaishi K., Kanda K., Kimizuka F., Kato I., Sobue K.;
RT	"Primary structure and functional expression of h-caldesmon complementary DNA.";
RL	Biochem. Biophys. Res. Commun. 164:503-511(1989).
RN	(3)
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORM BRAIN L-CAD).
RC	TISSUE=Brain;
RX	MEDLINE=91093148; PubMed=1824698;
RA	Hayaishi K., Fujio Y., Kato I., Sobue K.;
RT	"Structural and functional relationships between h- and l-caldesmons.";
RL	J. Biol. Chem. 266:355-361(1991).
RN	(4)
RP	SEQUENCE FROM N.A. (GIZZARD H-CAD; BRAIN L-CAD AND GIZZARD L-CAD).

RX MEDLINE=94071934; PubMed=8250919;
 RA Haruna M., Hayashi K., Yano H., Takeuchi O., Sobue K.;
 RT "Common structural and expressional properties of vertebrate
 RT caldesmon genes.";
 RL Biochem. Biophys. Res. Commun. 197;145-153(1992).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORM GIZZARD L-CAD).
 RC TISSUE=Gizzard;
 RX MEDLINE=92042686; PubMed=1939602;
 RA Bryan J., Lee R.;
 RT "Sequence of an avian non-muscle caldesmon.";
 RL J. Muscle Res. Cell Motil. 12;372-375(1991).
 RN [6]
 RP SEQUENCE OF 1-15 FROM N.A. (ISOFORM BRAIN L-CAD).
 RC TISSUE=Brain;
 RX MEDLINE=94271210; PubMed=8002994;
 RA Yano H., Hayashi K., Haruna M., Sobue K.;
 RT "Identification of two distinct promoters in the chicken caldesmon
 RT gene.";
 RL Biochem. Biophys. Res. Commun. 201;618-626(1994).
 RN [7]
 RP SEQUENCE OF 451-756 FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Gizzard;
 RX MEDLINE=89273666; PubMed=2730665;
 RA Hayashi K., Yamada S., Kanda K., Kimizuka F., Kato I., Sobue K.;
 RT "35 kDa fragment of h-caldesmon conserves two consensus sequences of
 RT the tropomyosin-binding domain in tropomyosin T.";
 RL Biochem. Biophys. Res. Commun. 161;38-45(1989).
 RN [8]
 RP SEQUENCE OF 498-525.
 RX MEDLINE=88293484; PubMed=3401222;
 RA Mornet D., Audenaert E., Derancourt J.;
 RT "Identification of a 15 kilodalton actin binding region on gizzard
 RT caldesmon probed by chemical cross-linking";
 RL Biochem. Biophys. Res. Commun. 154;564-571(1988).
 RN [9]
 RP PHOSPHORYLATION SITES.
 RX MEDLINE=92041815; PubMed=1939059;
 RA Mak A.S., Carpenter M., Smillie L.B., Wang J.H.;
 RT "Phosphorylation of caldesmon by p34cdc2 kinase. Identification of
 RT phosphorylation sites.";
 RL J. Biol. Chem. 266;119971-19975(1991).
 RN [10]
 RP PHOSPHORYLATION OF TYR-27 AND TYR-165.
 RX MEDLINE=20026923; PubMed=10559276;
 RA Wang Z., Danielssen A.J., Mahlie N.J., McMann M.J.;
 RT "Tyrosine phosphorylation of caldesmon is required for binding to the
 RT Shc-Gb2 complex.";
 RL J. Biol. Chem. 274;33807-33813(1999).
 CC -1- FUNCTION: ACTIN- AND MYOSIN-BINDING PROTEIN IMPLICATED IN THE
 CC REGULATION OF ACTOMYOSIN INTERACTIONS IN SMOOTH MUSCLE AND
 CC NONMUSCLE CELLS (COULD ACT AS A BRIDGE BETWEEN MYOSIN AND ACTIN
 CC FILAMENTS). STIMULATES ACTIN BINDING OF TROPOMYOSIN WHICH
 CC INCREASES THE STABILIZATION OF ACTIN FILAMENT STRUCTURE. IN MUSCLE
 CC TISSUES, INHIBITS THE ACTOMYOSIN ATPASE BY BINDING TO F-ACTIN.
 CC THIS INHIBITION IS ATTENUATED BY CALCIUM-CALMODULIN AND IS
 CC POTENTIATED BY TROPOMYOSIN. INTERACTS WITH ACTIN, MYOSIN, TWO
 CC MOLECULES OF TROPOMYOSIN AND WITH CALMODULIN. ALSO PLAY AN
 CC ESSENTIAL ROLE DURING CELLULAR MITOSIS AND RECEPTOR CAPING.
 CC SUBCELLULAR LOCATION: ON THIN FILAMENTS IN SMOOTH MUSCLE AND ON
 CC STRESS FIBERS IN FIBROBLASTS (NONMUSCLE) (BY SIMILARITY).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Name=Gizzard h-cad;
 CC IsoId=p12957-1; Sequence=Displayed;
 CC Name=Brain l-cad;
 CC IsoId=p12957-2; Sequence=vsp_004152, vsp_004153;
 CC Name=Gizzard l-cad;
 CC IsoId=p12957-3; Sequence=vsp_004153;
 CC -1- TISSUE SPECIFICITY: HIGH-MOLECULAR-WEIGHT CALDESMON (H-CALDESMON)
 CC IS PREDOMINANTLY EXPRESSED IN SMOOTH MUSCLES, WHEREAS LOW-
 CC MOLECULAR-WEIGHT CALDESMON (L-CALDESMON) IS WIDELY DISTRIBUTED IN
 CC NON-MUSCLE TISSUES AND CELLS NOT EXPRESSED IN SKELETAL MUSCLE OR

CC HEART.
CC -1- DOMAIN: THE N-TERMINAL PART SEEMS TO BE A MYOSIN/CALMODULIN-
CC BINDING DOMAIN, AND THE C-TERMINAL A TROPOMYOSIN/ACTIN/CALMODULIN-
CC BINDING DOMAIN. THESE TWO DOMAINS ARE SEPARATED BY A CENTRAL
CC HELICAL REGION IN THE MUSCLE FORMS.
CC -1- PTM: Phosphorylation in non-muscle cells. Phosphorylation by CDC2
CC during mitosis causes caldesmon to dissociate from microfilaments.
CC Phosphorylation reduces caldesmon binding to Actin, Myosin, and
CC Calmodulin as well as its inhibition of actomyosin ATPase
CC activity. Phosphorylation also occurs in both quiescent and
CC dividing smooth muscle cells with similar effects on the
CC interaction with Actin and Calmodulin and on microfilaments
CC reorganization (By similarity).
CC -1- SIMILARITY: BELONGS TO THE CALDESOMON FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: J04968; AAA49067.1; -.
CC EMBL: D17648; BAA04539.1; JOINED.
CC EMBL: D17634; BAA04539.1; JOINED.
CC EMBL: D17637; BAA04539.1; JOINED.
CC EMBL: D17638; BAA04539.1; JOINED.
CC EMBL: D17639; BAA04539.1; JOINED.
CC EMBL: D17640; BAA04539.1; JOINED.
CC EMBL: D17641; BAA04539.1; JOINED.
CC EMBL: D17642; BAA04539.1; JOINED.
CC EMBL: D17643; BAA04539.1; JOINED.
CC EMBL: D17644; BAA04539.1; JOINED.
CC EMBL: D17645; BAA04539.1; JOINED.
CC EMBL: D17646; BAA04539.1; JOINED.
CC EMBL: M28417; AAA48810.1; -.
CC EMBL: M06620; AAA48936.1; -.
CC EMBL: D17648; BAA04538.1; -.
CC EMBL: D17635; BAA04538.1; JOINED.
CC EMBL: D17636; BAA04538.1; JOINED.
CC EMBL: D17637; BAA04538.1; JOINED.
CC EMBL: D17639; BAA04538.1; JOINED.
CC EMBL: D17640; BAA04538.1; JOINED.
CC EMBL: D17641; BAA04538.1; JOINED.
CC EMBL: D17642; BAA04538.1; JOINED.
CC EMBL: D17643; BAA04538.1; JOINED.
CC EMBL: D17644; BAA04538.1; JOINED.
CC EMBL: D17645; BAA04538.1; JOINED.
CC EMBL: D17646; BAA04538.1; JOINED.
CC EMBL: D17647; BAA04538.1; JOINED.
CC EMBL: D17648; BAA04540.1; -.
CC EMBL: D17634; BAA04540.1; JOINED.
CC EMBL: D17636; BAA04540.1; JOINED.
CC EMBL: D17637; BAA04540.1; JOINED.
CC EMBL: D17639; BAA04540.1; JOINED.
CC EMBL: D17640; BAA04540.1; JOINED.
CC EMBL: D17641; BAA04540.1; JOINED.
CC EMBL: D17642; BAA04540.1; JOINED.
CC EMBL: D17643; BAA04540.1; JOINED.
CC EMBL: D17644; BAA04540.1; JOINED.
CC EMBL: D17645; BAA04540.1; JOINED.
CC EMBL: D17646; BAA04540.1; JOINED.
CC EMBL: D17647; BAA04540.1; JOINED.
CC EMBL: D17648; BAA04540.1; JOINED.
CC EMBL: M59762; AAA48649.1; -.
CC EMBL: D17552; BAA04490.1; -.
CC EMBL: M26684; AAA48811.1; -.
CC PIR: A33430; A33430.
CC HSP: P05412; IFOS.
CC InterPro: IPR006017; Caldesmon.
CC InterPro: IPR006018; Caldesmon_LSP.
CC Pfam: PF02029; Caldesmon; 1.

DR PRINTS; PRO1076; CALDESOMON.
KW Muscle protein; Actin-binding; Calmodulin-binding; phosphorylation;
KW Repeat; Alternative splicing. MYOSIN AND CALMODULIN-BINDING.
FT DOMAIN 26 199 10 X 13 AA APPROXIMATE TANDEM REPEATS.
FT FT 251 390 1. 1.
FT REPEAT 251 265 1. 1.
FT REPEAT 266 278 2. 2.
FT REPEAT 279 291 3. 3.
FT REPEAT 294 306 4. 4.
FT REPEAT 309 321 5. 5.
FT REPEAT 324 336 6. 6.
FT REPEAT 337 349 7. 7.
FT REPEAT 350 362 8. 8.
FT REPEAT 363 375 9. 9.
FT REPEAT 378 390 10. 10.
FT DOMAIN 523 580 TROPOMYOSIN-BINDING (POTENTIAL).
FT FT 622 632 TROPOMYOSIN-BINDING (POTENTIAL).
Query Match 20.5% Score 182.5; DB 1; Length 771;
Best Local Similarity 33.3%; Pred. No. 0.017; Mismatches 71; Indels 23; Gaps 6;
Matches 61; Conservative 28; Mismatches 71; Indels 23; Gaps 6;
QY 5 REAAEEBAKRRADAKKQKPKRAKGVPGHATPD-----KKENDAKSSDSVGEET 57
DB 171 RQDEEECKKKEKXSEEEKPK-----EVPTEBNQVDVAVEKSTDKKEVEVETKTLAVNEN 225
QY 58 LPSPSLPEKKVAE-AEKVVEAKKKAEDQKEDRRNYPTTYTLEFIASDVEYKA 116
DB 226 DTNAMLBEQSIITPAADKKEKKEKKEKLEAEERKLEAEKKAEEKKKAEEKKKA 285
QY 117 ELELVKEBAK---BPRNEKXKQAKAEVSKAEATRLKTKTRKKAEEBAK---RKA 169
DB 286 E---ERRAKAEEKRAAEERAKAEERAKAEERAKAEERAKAEERAKAEERAKA 342
QY 170 AEE 172
DB 343 AEE 345
RESULT 3
IDF1_SCHPO STANDARD; PRT; 1403 AA.
IDF1_SCHPO
AC 010475;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable eukaryotic initiation factor C17C9.03.
GN SPAC17C9.03.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooke K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeli C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Gymnopoulos B.,
RA Waltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,

```

RA Eger P., Zitzmann W., Wedler H., Manburt R., Purrelle B.,
RA Goffeau A., Cadieu E., Drenth S., Gloux S., Lelaure V., Motlier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revelante J.L., Moreno S., Armstrong J., Foreburg S.L.,
RA Cernutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakowski G.V., Usery D., Barrett B.G., Nurse P.,
RT "The genome sequence of Schizosaccharomyces pombe.",
RL Nature 415:871-880(2002).
-1- SIMILARITY: TO YEAST EUKARYOTIC INITIATION FACTOR 4F SUBUNIT'S P130
AND P150.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z73099; CAA97349.1; -.
DR PIR; T11583; T11583.
DR GenBank; SPAC109.03; -.
DR InterPro; IPR003890; IPR_elf4G.
DR Pfam; PF02854; MIF4G; 1.
DR SMART; SM00543; MIF4G; 1.
KM Hypothetical protein; Initiation factor; Protein biosynthesis;
KM RNA-binding.
SQ
SEQUENCE 1403 AA; 154034 MW; 0317EE65BE2A1B53 CRC64;

Query Match 20.4%; Score 181.5; DB 1; Length 1403;
Best Local Similarity 31.3%; Pred. No. 0.03;
Matches 56; Conservative 35; Mismatches 57; Indels 31; Gaps 7;

QY 5 REEAREEKKRRADAKEOGKPKRKARQVPGELATDPDKKENDKSSDSVGEETLSPSLK 64
DB 562 KREAEQAKREAEBAEKAKREAEBAE-EKAKREAEENAKREAEBAEKRA----- 613
QY 65 PEKKAEEAEKKEE-BAKKAEDQKEEDRRNPPTNTKYLEIAESDVEVKKAELELVEK 123
DB 614 EEKAKREAEBAEKAKREAEBAEKAKREAEBAEKAKR-----EAEBAKREAEBAEKAKR 659
QY 124 EAKEPRNEEKKYQKQAAEV-ESKKAATRLKTKTD-----KKAAEEAKRKAAEDKVK 176
DB 660 EAEENAKREAEBAEKAKREAEENAKREAE--EKVKRETEENAKRKAAEEGKREADKNPEIK 716

RESULT 4
NFH_MOUSE STANDARD; PRT; 1087 AA.
ID_NFH_MOUSE
AC P19246; O61959;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neurofilament triplet H protein (200 kDa neurofilament protein)
DE (Neurofilament heavy polypeptide) (NF-H).
GN NFH OR NFH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN (1)
RX MEDLINE=69121513; PubMed=3220257;
RA Julien J.-P., Cote F., Beaudet L., Sidky M., Flavell D., Grosveld F.,
RA Mushynski W.;
RT "Sequence and structure of the mouse gene coding for the largest
RT neurofilament subunit.";
RL Gene 68:307-314(1988).
RN (2)
RX SEQUENCE FROM N.A.
RX MEDLINE=89089138; PubMed=3145094;

```

RA	Shneiderman P.S., Carden M.J., Lee J.F., Lazzarini R.A.;
RT	"The structure of the largest murine neurofilament protein (NF-H) as revealed by cDNA and genomic sequences";
RL	Brain Res. 464:217-231(1988).
RN	[3]
RP	SEQUENCE FROM N.A.
RC	STRAIN=Swiss Webster; TISSUE=Brain;
RA	Carden M.J.;
RL	Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC	-I- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC	NF-H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC	NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT SUBSERVED BY THE TWO SMALLER NF PROTEINS.
CC	-I- PM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P. NFH IS PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE OF AXONAL CALIBER.
CC	-I- PM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC	-I- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC	-I- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 534 TO 716 AND IS SHORTER DUE TO FRAMESHIFTS.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; M24496; AAA39813.1; .
DR	EMBL; M23349; AAA39813.1; JOINED.
DR	EMBL; M24494; AAA39813.1; JOINED.
DR	EMBL; M24495; AAA39813.1; JOINED.
DR	EMBL; M35131; AAA39809.1; ALT_FRAME.
DR	EMBL; Z31012; CAA83229.1; .
DR	PIR; JT0368; QPM5H.
DR	MGI: 97309; Neth.
DR	InterPro: IPR01664; IF.
DR	Pfam; PF00038; Filament; 1.
DR	PROSITE; PS00226; IF; 1.
KW	Intermediate filament; Coiled coil; Neutrone; Phosphorylation; Repeat.
KW	DOMAIN
FT	DOMAIN
FT	DOMAIN
FT	DOMAIN
FT	DOMAIN
FT	DOMAIN
FT	DOMAIN
FT	DOMAIN
FT	DOMAIN
FT	DOMAIN
FT	DOMAIN
FT	DOMAIN
FT	DOMAIN
FT	CONFLICT
FT	CONFLICT
FT	CONFLICT
FT	CONFLICT
FT	CONFLICT
FT	CONFLICT
FT	CONFLICT
SO	SEQUENCE
Query Match	19.9%; Score 177; DB 1; Length 1087;
Best Local Similarity	33.7%; Pred. No. 0.043;

Matches 65; Conservative 21; Mismatches 73; Indels 34; Gaps 9;

QY 5 REBAEEKRRADAKGQKGRAGVPGELATPDKKENDAKSSDSSVGEE-TLPSPL 63
 Db 859 KEKASPEKEEKTSEKAPKKE-----EKSVEKEVKAKEPPKVEBEKTLPT- 909
 QY 64 KPEKVAEAKKVEBAKKAEDQKEEDRRNYPNTYKTLLEIAESDVEVK---AELE 119
 Db 910 KTEAK-----ESKQDEAPKEAPKVEEKEKPTPEKXOSTAEAKKEEGEKKNVASEE 965
 QY 120 L-----VKEAKEPRNEKVKQAKAEVSKA---EATRLKTKTDKKAEEAK---- 166
 Db 966 TPALKVGEAKAP--KEKETITTKTEADDTAKESKPTETEKPKKEEMPAPAEKDTKE 1022
 QY 167 RKAABEDVKKEP 179
 Db 1023 EKTESRKPEEK 1035

RESULT 5

NFM_RABIT STANDARD; PRT; 644 AA.
 AC PS4938;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Neurofilament triplet M protein (160 kDa neurofilament protein)
 DE (Neurofilament medium polypeptide) (NF-M) (Fragment).
 GN NEF3 OR NEFM OR NFM.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Heart;
 RX MEDLINE=97055255; PubMed=8899542;
 RA Vitadello M., Vettore S., Lamar E., Chien K.R., Gorza L.;
 RT "Neurofilament M mRNA is expressed in conduction system myocytes of
 the developing and adult rabbit heart."
 RL J. Mol. Cell. Cardiol. 28:1833-1844(1996).
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
 AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 CC -1- PPM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P. NFM IS
 PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
 THOUGHT THAT PHOSPHORYLATION OF NFM RESULTS IN THE FORMATION OF
 INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
 OF AXONAL CALIBER.
 CC -1- PPM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
 OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H). THE
 LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
 COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z47378; CAAB7454.1; -;
 DR PIR; S55395; S55395.
 DR InterPro; IPR001664; IP.
 DR Pfam; PFO0038; Pfam; 1.
 DR PROSITE; PS00226; IP; 1.
 KW Intermediate filament; Coiled coil; Neurone; Phosphorylation;
 KW Glycoprotein.
 FT NON_TER 1 197 ROD.
 FT DOMAIN <1 198 643 TAIL.
 FT DOMAIN <1 33 COIL 1B.
 FT

FT DOMAIN 34 50 LINKER 12.
 FT DOMAIN 51 72 COIL 2A.
 FT DOMAIN 73 76 LINKER 2.
 FT DOMAIN 77 197 COIL 2B.
 FT CARBOHD 217 217 O-LINKED (GLNAC) (BY SIMILARITY).
 SQ SEQUENCE 644 AA; 72450 MW; 030FDAA622889678 CRC64;

Query Match 19.9%; Score 176.5; DB 1; Length 644;
 Best Local Similarity 28.3%; Pred. No. 0.029;
 Matches 63; Conservative 30; Mismatches 85; Indels 37; Gaps 7;

QY 1 IKTDREBAEBAKRRADAKGQKGRAGVPGELATPD-KKENDAKSSDSSVGEE-TLP 59
 Db 268 VSVKEEKEBAEBAEKEGEQAEVEVAAKKS-PYKATTPETKEEKEEKEEKEEKEE 326
 QY 60 SPSLK-----PEKVAEAKVVEA---KKRAEDQKE-DRNRY 94
 Db 327 DEGVKSDQAEKGESKESKNEGEQEGETAEAEVEBAKEKTEESSEVAKEE 386
 QY 95 PTNTYKTLLEIAESDVEVKKAELELVKEAKEPRNEKVKQAKAEVSKA---EATRL 152
 Db 387 PVTEAKVGEKAKSPKPSVVEVKRAEATAGKECKEKEEVEEKKAAKESPKKE 446
 QY 153 KIK-----TDKKAEEBAKRAABEDVKKEP 179
 Db 447 KVEKKEKPKQVPRKKAESPVEKAEAEATTKP 481

RESULT 6

MNN4 YEAST STANDARD; PRT; 1178 AA.
 AC P36044; P36043; P89095;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE MNN4 Protein.
 GN MNN4 OR YKL200C/YKL201C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288C;
 RX MEDLINE=97175967; PubMed=9023541;
 RA Odani T., Shizuma Y.-I., Tanaka A., Jigami Y.;
 RT "Cloning and analysis of the MNN4 gene required for phosphorylation
 of N-linked oligosaccharides in Saccharomyces cerevisiae."
 RT Glycobiology 6:805-810(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Mata e Silva A., Bossier P., Vilela C., Fernandes L., Soares H.,
 RA Guerreiro P., Rodrigues-Pousada C.;
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: MAY FUNCTION AS A POSITIVE REGULATOR FOR
 NANNOSYLPHOSPHATE TRANSFER IN BOTH THE CORE AND OUTER CHAIN
 PORTIONS OF N-LINKED OLIGOSACCHARIDES.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
 CC -1- SIMILARITY: TO YEAST YJ061W.
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO
 FRAMESHIFTS, ONE OF WHICH PRODUCES TWO SEPARATE ORFS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D83006; BA11676.1; -;
 DR EMBL; Z28201; CAAB2046.1; -;

[illegible]

```

DT 01-MUG-1990 (Rel. 15, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurofilament triplet H protein (200 kDa neurofilament protein)
DE (Neurofilament heavy polypeptide) (NF-H) (Fragment).
DE NEFH OR NFH.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN NM_011111
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA MEDLINE=89065087; PubMed=3143306;
RA Breen K.C., Robinson P.A., Wion D., Anderson B.H.;
RT "Partial sequence of the rat heavy neurofilament polypeptide (NF-H).
RT Identification of putative phosphorylation sites."
RN FEBS Lett. 241:213-218(1988).
[2]
RP SEQUENCE OF 37-831 FROM N.A.
RA MEDLINE=88309090; PubMed=2457365;
RA Dautigny A., Pham-Dinh D., Rousset C., Felix J.M., Nussbaum J.L.,
RA Jolles P.;
RT "The large neurofilament subunit (NF-H) of the rat: cDNA cloning and
RT in situ detection."
RN Biochem. Biophys. Res. Commun. 154:1099-1106(1988).
[3]
RP SEQUENCE OF 1-89 AND 243-313 FROM N.A.
RA MEDLINE=87080760; PubMed=2878828;
RA Robinson P.A., Wion D., Anderson B.H.;
RT "Isolation of a cDNA for the rat heavy neurofilament polypeptide
RT (NF-H).";
RN FEBS Lett. 209:203-205(1986).
[4]
RP SEQUENCE OF 318-831 FROM N.A.
RA MEDLINE=89184647; PubMed=2928342;
RA Lieberburg I., Spinner N., Snyder S., Anderson J., Goldgaber D.,
RA Smulowitz M., Carroll Z., Emanuel B.S., Breitner J., Rubin L.;
RT "Cloning of a cDNA encoding the rat high molecular weight
RT neurofilament peptide (NF-H): developmental and tissue expression in
RT the rat, and mapping of its human homologue to chromosomes 1 and
RT 22.";
RN Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).
[5]
RP FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
RP AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
RP NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
RP SUBSERVED BY THE TWO SMALLER NF PROTEINS.
[6]
RP -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P. NFH IS
RP PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
RP THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
RP INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
RP OF AXONAL CALIBER.
[7]
RP -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
RP OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H). THE
RP LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
RP COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
[8]
RP -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
RP -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 783
RP ONWARD AND IS LONGER DUE TO A FRAMESHIFT.
[9]
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; M37227; AAA41693.1; ALT_FRAME.
DR EMBL; X13804; CAA32038.1; ALT_FRAME.
DR EMBL; M21964; AAA41695.1; -.
DR EMBL; J04517; AAA41692.1; -.
DR InterPro; IPR001664; IP.

```

DR	Pfam: PF00038; filament; 1.
DR	PROSITE; PS00226; IF; 1.
KW	Intermediate filament; Coiled coil; Neurone; Phosphorylation; Repeat. NON_TER 1
FT	DOMAIN 276 641 51 X 3 AA TANDEM REPEATS OF K-S-P.
FT	CONFLICT 164 164 L -> I (IN REF. 2).
FT	CONFLICT 185 185 L -> S (IN REF. 2).
FT	CONFLICT 193 193 L -> T (IN REF. 2).
FT	CONFLICT 199 199 M -> T (IN REF. 2).
FT	CONFLICT 346 346 K -> N (IN REF. 1).
FT	CONFLICT 373 373 A -> V (IN REF. 1 AND 4).
FT	CONFLICT 482 482 G -> E (IN REF. 2 AND 4).
FT	CONFLICT 485 485 P -> S (IN REF. 2).
FT	CONFLICT 570 571 RK -> KE (IN REF. 2 AND 4).
FT	CONFLICT 591 591 P -> T (IN REF. 2 AND 4).
FT	CONFLICT 727 727 A -> V (IN REF. 4).
FT	CONFLICT 757 759 AAP -> GST (IN REF. 4).
FT	CONFLICT 769 769 T -> L (IN REF. 2).
FT	CONFLICT 775 775 R -> P (IN REF. 2 AND 4).
SQ	SEQUENCE 831 AA; 89486 MW; 1B0973C3F13EF768 CRC64;
Query Match	19.4%; Score 172.5; DB 1; Length 831;
Best Local Similarity	29.1%; Pred. No. 0.056;
Matches	60; Conservative 35; Mismatches 68; Indels 43; Gaps 7
QY	11 EAKRRADAEQGCKPGRAGVPGELATPDKKEN-----DAKSSD---SSVGEEF 57
DB	494 EAKSPAEKAPPAEAKSPAEAKSPVEVKSEPKAKSPVKEGAKSLAEKSPAEKSPVKEI 553
QY	58 LPSPBLK-----PEKKVAEAERKV-----BEAKKKAEDQAEDERNRP 95
DB	554 KPRAEVKSPPEAKSPMRKEAASPEAKTLDVKSPEAKPPAKEAKRPAIRISPEOVKSPA 613
QY	96 INTVKTLELAESDVVEVKALELVKEEKAPRNBEVKQAKAEVSKKAAATLREIKX 155
DB	614 KEPAKSPPEKE-----ETRTKEVAAKKEGVKSPVEVAKAPEPKVYEETKTATTEKTVK 667
QY	156 TDRK-KAEER-KRKAAEDDKVERP 179
DB	668 ESKKDAPKEAQKPRAEKEPELTERP 693
RESULT 8	
ID	IF2P_HUMAN STANDARD; PRT; 1220 AA.
AC	O60841; O95805; Q9UF81; Q9UN7;
DT	15-DEC-1998 (Rel. 37, Created)
DT	16-OCT-2001 (Rel. 40, Last sequence update)
DT	28-FEB-2003 (Rel. 41, Last annotation update)
DE	Translation initiation factor IF-2.
GN	IF2 OR KIAA0741.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Brain;
RX	MEDLINE=99087487; PubMed=9872452;
RA	Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RA	Tanaka A., Kotani H., Nomura N., Ohara O.;
RT	"Prediction of the coding sequences of unidentified human genes. XI.
RT	The complete sequences of 100 new cDNA clones from brain which code
RT	for large proteins in vitro.";
RL	DNA Res. 5:277-286(1998).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Cervical carcinoma;
RX	MEDLINE=99362399; PubMed=10432305;
RA	Wilson S.A., Stietro-Vazquez C., Edwards N.J., Iourin O., Byles E.D.,
RA	Kocsozpoulou E., Adamson C.S., Kingsman S.M., Kingeman A.J.,
RT	Martin-Rendon E.;
RT	"Cloning and characterization of hIF2, a human homologue of bacterial

RT				translation initiation factor 2 and its interaction with HIV-1 matrix.";
RL				Biochem. J. 342:97-103(1999).
RN	(3)			
RP				SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC				TISSUE=Testis;
RX	MEDLINE=99218282; PubMed=10200264;			
RA	Lee J.H., Choi S.K., Koli-Meck A., Burley S.K., Dever T.E.;			
RA	"Universal conservation in translation initiation revealed by human			
RT	and archaeal homologs of bacterial translation initiation factor			
RT	IF2.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 96:4342-4347(1999).			
RN	[4]			
RP	SEQUENCE OF 89-1220 FROM N.A.			
RC	TISSUE=Testis;			
RA	Koehlerer K., Beyer A., Mewes H.-W., Gassenhuber J., Wiemann S.;			
RL	Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.			
RN	[5]			
RP	SEQUENCE OF 833-1220 FROM N.A.			
RX	MEDLINE=21064499; PubMed=1124703;			
RA	Stanchi F., Bertocco E., Toppo S., Diognardi R., Simonati B.,			
RA	Carnata N., Zimbellio R., Lanfranchi G., Valle G.;			
RT	"Characterization of 16 novel human genes showing high similarity to			
RT	yeast sequences.";			
RL	Yeast 18:69-80(2001).			
CC	-I- FUNCTION: FUNCTION IN GENERAL TRANSLATION INITIATION BY PROMOTING			
CC	THE BINDING OF THE FORMYL METHIONINE-TRNA TO RIBOSOMES. SEEMS TO			
CC	FUNCTION ALONG WITH EIF-2 (BY SIMILARITY).			
CC	-I- SIMILARITY: BELONGS TO THE IF-2 FAMILY.			
CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on ways			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/ or_send_an_email_to_license@isb-sib.ch).			
CC				
DR	EMBL; AB018284; BAA34461.1; -.			
DR	EMBL; AJ006776; CAB44357.1; -.			
DR	EMBL; AF078035; AAD16006.1; -.			
DR	EMBL; AL133563; CAB63717.1; -.			
DR	EMBL; AJ006412; CAA07018.1; -.			
DR	PIR; T43483; T43483.			
DR	MIM; 606086; -.			
DR	GO; GO:0003743; F:translation initiation factor activity; NAS.			
DR	GO; GO:0006446; P:regulation of translational initiation; NAS.			
DR	InterPro; IPR000795; EF_GTP_D2.			
DR	InterPro; IPR004161; EFTU_D2.			
DR	InterPro; IPR000178; IF2.			
DR	InterPro; IPR005225; Small_GTP.			
DR	pfam; PF00009; GTP_EFTU_1.			
DR	pfam; PF03144; GTP_EFTU_D2_1.			
DR	PRINTS; PR00315; ETONGATNFCT.			
DR	TIGRFAMs; TIGR00231; small_GTP_1.			
DR	PROSITE; PS01176; IF2; FALSE_NEG.			
KW	Initiation factor; Protein biosynthesis; GTP-binding.			
FT	NP_BIND	638	645	GTP (BY SIMILARITY).
FT	DOMAIN	39	50	POLY-LYS.
FT	DOMAIN	94	99	POLY-LYS.
FT	DOMAIN	138	142	POLY-ASP.
FT	DOMAIN	313	322	POLY-LYS.
FT	DOMAIN	353	356	POLY-GLU.
FT	DOMAIN	361	364	POLY-GLU.
FT	DOMAIN	491	496	POLY-GLU.
FT	DOMAIN	529	567	ASF/GLU-RICH (ACIDIC).
FT	DOMAIN	640		V-G: LOSS OF ACTIVITY IN VIVO. RETAINS FULL ACTIVITY IN VITRO.
FT	MUTAGEN	706		H-E: LOSS OF ACTIVITY; BOTH IN VIVO AND IN VITRO.
FT	MUTAGEN	706		H-Q: LOSS OF ACTIVITY IN VIVO. PARTIAL ACTIVITY IN VITRO.
FT	MUTAGEN	706		D-N: LOSS OF ACTIVITY; BOTH IN VIVO AND
FT	MUTAGEN	759	759	

	FT	CONFLICT	IN VITRO
FT	64	64	E -> G (IN REF. 1).
FT	92	92	T -> I (IN REF. 2).
FT	180	180	I -> M (IN REF. 3).
FT	256	256	K -> R (IN REF. 3).
FT	522	522	T -> K (IN REF. 1).
FT	549	549	E -> V (IN REF. 3).
FT	669	669	G -> W (IN REF. 3).
FT	894	894	E -> K (IN REF. 2).
SEQ	SEQUENCE	1220 AA; 138799 MW; 010C15DE475E5B8 CRC64;	
Query Match	19.3%;	Score 112; DB 1; Length 1220;	
Best Local Similarity	28.8%;	Pred. No. 0.082;	
Matches	57; Conservative	40; Mismatches 69; Indels 32; Gaps 7	
QY	1	IKTDEEAEAEKRRADAEQSGPKGRARGVGELATPDKKENDAKSSGSSVGSEET---	57
DB	225	IKTVAQKAEEKERERKKDEKAKLR-KLKEKELETKQKQSKOKESQRFEEETVKS	283
QY	58	-----LPSPSLPEKKVA-----EAEKTYEAAKK--AEQKEEDRNRYPTN-TYKT	101
DB	284	KYTVDTGVIPAEBEKAETPTTAADNDEGKKKKKKKKGGKEEKEKEKKKSPKATYKA	343
QY	102	LELEIAESDVEYKKALELVKEEAKPRNEEYKQAKAVESSKAEATPLEKTKTDKKK-	160
DB	344	MQEALA-----KLKEEERQKREEBEIRIKRLBELAKRKEEBERLEQEKERKKQ	392
QY	161	AEEAKRAAEEDYKKEK	178
DB	393	KEKERKERLKEEGKLLTK	410

```

RESULT 9
IF2_AQUAE
ID IF2_AQUAE STANDARD: PRT: 805 AA.
AC 067825;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Translation initiation factor IF-2.
UN INP3 OR AQ.2032.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxId=633163;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RA MEDLINE=98196666; PubMed=9537320;
RA Decker E.G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swenson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus."
RL Nature 392:353-358(1998).
CC -!- FUNCTION: One of the essential components for the initiation of
CC protein synthesis. Protects formylmethionyl-tRNA from spontaneous
CC hydrolysis and promotes its binding to the 30S ribosomal subunits.
CC Also involved in the hydrolysis of GTP during the formation of the
CC 70S ribosomal complex (by similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE IF-2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AE000769; AAC07794.1; -.
DR PIR, E70474; E70474.
DR HAMAP, MF_00100; -; 1.

```

Query Match	Score	DB	Length
Best Local Similarity	29.08%	Pred. No. 0.06;	805;
Matches 56; Conservative	33; Mismatches	53; Indels	51; Gaps

```

QY 4 DREAEAEARRKRAADAKGGRKAGKGVPELALTPPKENDAKSDSSVVEETLPPSPL 63
Db 61 EEEEEEEVVTQAAQ-----PAEV--EEKKEEKK--EEVIEEVEEK-- 100
QY 64 KPEKVAEAEKKVEAEKAKKADQEDRRNYPNTYKTL-----LEIAESDVEKKAE 117
Db 101 KPEVIVEIEIEKKKEEKEEKKERK-----KSVEELIKEIIEKKKEKKKVE 149
QY 118 -----LELYKEAKKERPNNEKKYQAAEVESSKAEKTRLEK---IKTRKKAEE 163
Db 150 KERKEEKRVVVEYKKEEKEEKKKEEKKPKIKMSKKEREIRMKLEHAVEKEKKQEK 209
QY 164 EAKKRAAEEDKVK 176
Db 210 REKEKKKKEEVK 222

RESULT 10
MST1_DROHY STANDARD. PRT: 344 AA.
AC Q08695:
AC 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Axoneme-associated protein ms101(1).
GN MST101(1).
OS Drosophila hydei (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7224;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE=Testis;
RX MEDLINE=94200512; PubMed=8150205;
RA Neesen J., Buenemann H., Heinlein U.A.;
RT "The Drosophila hydei gene dhms101(1) encodes a testis-specific,
RT repetitive, axoneme-associated protein with differential abundance in
RT Y chromosomal deletion mutant flies.";
RL Dev. Biol. 162:414-425(1994).
CC -1- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL. IT IS
CC ASSOCIATED WITH AXONEMAL STRUCTURES.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
CC -1- TISSUE SPECIFICITY: TESTIS. LOCATED IN SPERMATOCYTES AND
CC SPERMATID BUNDLES.
CC -1- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.

```

[illegible]

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, X73481, CAA51876.1; -.
DR PIR, S51364; S51364.
DR FlyBase, FBgn0020733; Dhyd\msc10(2).
KW Sperm; Repeat; Multigene family; Polyomorphism.
FT DOMAIN             332      1268
                        59 x 16 AA APPROXIMATE TANDEM REPEATS OF
                        [KR]-K-X-C-X-X-A-K-X-X-K-X-X-E.
SQ
SEQUENCE           1391 AA;   159000 MW;   1B2A366F30F48878 CRC64;

Query Match              19.1%; Score 169.5; DB 1; Length 1391;
Best Local Similarity    33.9%; Pred. No. 0.12;
Matches     65; Conservative 28; Mismatches 64; Indels 35; Gaps 9

QY      5 REEAEEFKRADAKE-OGKKGAKAGVPELATPDCKENDAK-----SSDSVGGEELT 58
DB      306 KEQAEEEEKJGVVAEVEKKCKEKALKKKCDLGRMKKEAEKKCAALAKQKDEBK 365
QY      59 PSPSLPKPKYAABKKVVEA---KKKAEDQKE-----EDRRNYPTNTYKTLEIAESD 110
DB      366 ACKELAKKKGADEKKCEBAANKKKLAEEKCKEAKANEK-----EAEX- 412
QY      111 VEVKAAE--LELVYEAKEPNEBEKVQAKAVESSKA-EATRLERIKTDREKKAEEAK- 166
DB      413 ---KKCEBAKKKEAPAEKKCKCELALNIKKAEBKKCKEAKKEAEKKCBELAKK 469
QY      167 -RKAABEDVKE 177
DB      470 IKKAEEKKCEE 481

RESULT 12
ATTRX_CAEEL STANDARD; PRT; 1359 AA.
AC Q9UTEO; O02061;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Transcriptional regulator ATRX homolog (X-linked nuclear protein-1).
OS XNP-1 OR B0041.7.
CN Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea;
OC Rhabdilitidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RS SEQUENCE FROM N.A., AND CHARACTERIZATION.
RA MEDLINE=99365296; Pubmed=10433961;
RX Villard L., Fontes M., Embank J.J.;
RT "Characterization of xnp-1, a Caenorhabditis elegans gene similar to
RL the human XNF/ATR-X gene.";
RL Gene 236:13-19(1999).
RP [2]
RS SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Fulton R., Wohlmann P.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES
CC GENE EXPRESSION BY AFFECTING CHROMATIN (Potential).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC -----
DR EMBL: AF134186; AAD55361.1; -
DR EMBL: AF000196; AAC24256.1; -
DR PIR: T34036; T34036.
DR WormPep; B0041.7; CEI7314.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR000330; SNF2_N.
DR Pfam; PF002771; helicase_C; 1.
DR Pfam; PF00176; SNF2_N; 1.
DR SMART; SM00487; DEXDc; 1.
DR SMART; SM00490; HELICc; 1.
DR DNA repair; Nuclear protein; ATP-binding; Helicase; ATP-binding.
FT NP_BIND 496 503
FT SITE 636 639
FT DOMAIN 67 70
FT DOMAIN 266 272
FT DOMAIN 276 281
FT DOMAIN 372 375
FT DOMAIN 603 608
FT DOMAIN 859 862
FT CONFLICT 479 479 C->F (IN REF. 2).
SQ SEQUENCE 1359 AA; 156191 MW; EB4342547D4F4E64 CRC64;

Query Match 19.0%; Score 169; DB 1; Length 1359;
Best Local Similarity 26.9%; Pred. No. 0.13;
Matches 54; Conservative 40; Mismatches 81; Indels 26; Gaps 5;

OY 2 KTDREAEAEARRADAEQCKPKGRKGVPGELATPDCKENDAKSSDSVGEETLPSPLK 61
DB 80 KKSRRKAKSESSESDSEEDRKKSKKKVQCKKESKKKRTTSSSEDESDDEREQK 139
OY 62 SLKPKKV-----AAEKVPERA-----KKKADQCKEDRRNYPNTTYTTL 102
DB 140 SKKSKSKTKTKOTSSSSSESESEERKVKSKKCKEKSVKRAETSESDERKPSKSKG 199
OY 103 ELEIASVEVYKAELELVKEAKEPRMEKVKYKAKAEVSEKKAATLEKIKTKRKKAE 162
DB 200 LKKAKKSESE--SSEDEKVKSKKKSKKKVKKSESEDEPAKPKTKTERKKSKTISE 256
OY 163 EEAK-RKAEDPKVE---KP 179
DB 257 ESSESEKSDSEEEKESSPKP 277

RESULT 13
1E68_HSVSA STANDARD; PRT; 407 AA.
ID 1E68_HSVSA Q01042;
AC 001042;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE Immediate-early protein.
GN 73 OR ECLF1.
OS Herpesvirus saimiri (strain 11).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OC NCBI_TaxID=10383;
RX MEDLINE=92333688; PubMed=1321287;
RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,
RA Newman C., Wiltmann S., Craxton M.A., Coleman H., Fleckenstein B.,
RA Honess R.W.;
RT "Primary structure of the herpesvirus saimiri genome.";
RL J. Virol. 66:5047-5058 (1992).
RP MEDLINE=92230228; PubMed=1314457;
RA Nicholas J., Cameron K.R., Coleman H., Newman C., Honess R.W.;
RT "Analysis of nucleotide sequence of the rightmost 43 kbp of
RT herpesvirus saimiri (HVS) L-DNA: general conservation of genetic
RT organization between HVS and Epstein-Barr virus.";
```

```
RL Virology 188:296-310 (1992).
CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUP TOGETHER HSV-1 AND
CC HSV-2 IE-68 (US1), ENV-1 65, ENV-4 (OR4), PRV RSP40, AND VZV 63.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL: X64346; CAA45696.1; -
DR EMBL: M86409; AAA46149.1; -
DR EMBL: S76368; AAB21116.1; -
DR Early protein.
FT DOMAIN 60 241
FT SEQUENCE 407 AA; 46617 MW; FPD399CA82CE136C CRC64;

Query Match 19.0%; Score 168.5; DB 1; Length 407;
Best Local Similarity 26.6%; Pred. No. 0.046;
Matches 46; Conservative 44; Mismatches 50; Indels 33; Gaps 5;

OY 5 REAEAEAEARRADAEQCKPKGRKGVPGELATPDCKENDAKSSDSVGEETLPSPLK 64
DB 54 QQAALTEQRRREVEVEEGEE--RERG-----EEERGGEGEGEGE-----E 95
OY 65 PEKVAAEAEKVEBAKKAEDQCKEDRRNYPNTTYTTLLEIASDYVYKAELELYKEE 124
DB 96 AEEBAEAEKEAEBAEBAEBAEBAE-----EAEBAEBAEBAEBAEBAE 141
OY 125 AKEPRNEKVKQAKAEVSEKKAATRLKIKTKRKAEEAEKRAEDPKVE 177
DB 142 EAEBAEBAEBAEBAEBAEBAEBAEBAEBAE--EAEBAEBAEBAEBAEBAE 193

RESULT 14
MAPB_HUMAN STANDARD; PRT; 2468 AA.
ID MAPB_HUMAN P46821;
AC P46821;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Microtubule-associated protein 1B (MAP 1B) [Contains: MAP1 light chain
DE LC1].
GN MAP1B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
RX MEDLINE=95104835; PubMed=7806212;
RA Lien L.-Y., Feener C., Fischbach N., Kunkel L.M.;
RT "Cloning of human microtubule-associated protein 1B and the
RT identification of a related gene on chromosome 15.";
RL Genomics 22:273-280 (1994).
CC -1- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.
CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES
CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST
CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS
CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN
CC STABILIZING MICROTUBULES.
CC -1- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
CC WITH MAP1A AND MAP1B PROTEINS.
CC -1- DOMAIN: Has a highly basic region with many copies of the sequence
CC KKEE and KKEI/V, repeated but not at fixed intervals, which is
CC responsible for the binding of MAP1B to microtubules.
CC -1- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated
CC from MAP1B by proteolytic processing. It is free to associate with
CC both MAP1A and MAP1B. It interacts with the amino-terminal region
CC of MAP1B (By similarity).
```


CC -1- SIMILARITY: TO MAP1A.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L06337; AAA18904.1; -
 CC Genew; HGNC:6836; MAP1B.
 CC MIM; 157129; -
 CC GO; GO:0005875; C:microtubule associated complex; TAS.
 CC InterPro; IPR000102; MAP1B_neuraxin.
 CC Pfam; PF00414; MAP1B_neuraxin; 10.
 CC PROSITE; PS00230; MAP1B_NEURAXIN; 6.
 CC Microtubules; Repeat; Phosphorylation.
 CC CHAIN ? 2468 MAP1 LIGHT CHAIN LC1.
 FT REPEAT 1878 1894 MAP1B 1.
 FT REPEAT 1895 1911 MAP1B 2.
 FT REPEAT 1912 1928 MAP1B 3.
 FT REPEAT 1929 1945 MAP1B 4.
 FT REPEAT 1946 1962 MAP1B 5.
 FT REPEAT 1963 1979 MAP1B 6.
 FT REPEAT 1997 2013 MAP1B 7.
 FT REPEAT 2014 2030 MAP1B 8.
 FT REPEAT 2031 2047 MAP1B 9.
 FT REPEAT 2048 2064 MAP1B 10.
 FT DOMAIN 589 790 LYS-RICH (HIGHLY BASIC, CONTAINS MANY
 FT KEE AND KEEI/V REPEATS).
 SQ SEQUENCE 2468 AA; 270618 MW; 540839C8DF09D61 CRC64;
 Query Match 19.0%; Score 168.5; DB 1; Length 2468;
 Best Local Similarity 29.3%; Pred. No. 0.22;
 Matches 63; Conservative 31; Mismatches 76; Indels 45; Gaps 7;
 QY 2 KTDREAEAEKRRADAKGQK-PKGRKRGVPGELATPDCKENDAKSDSSVGEETLPS 60
 DB 555 KSVKESKEETPEYTKVNHVKEPKVESKEKVMKKDKPVETKTPSTKEKVSKEEPS 614
 QY 61 P-----SLKPEKVAEAKKVEAKKKAEDOKEDRRNYPTNTY 99
 DB 615 PVKAEVAKQKATDVKPKAKKEKTKKVKPEDEKKEKPKVEAKKEK-----TPIK 670.
 QY 100 KTLLELAESVVEYKAELELVKEBAKPRNE-----EKVKQAEVSKKAE 147
 DB 671 KEKPKKEEVEKVEYK--EIKKEEKEPKKEKVEKTEPPKEKVEKKEKVEKKEKE 727
 QY 148 ATR-LEKIKTDKKAEE--EAKKKAEDVKYKX 178
 DB 728 PKKEIKKLPKDAKKSSTPLSEAKRPALKPVPKK 762
 RESULT 15
 NFM_RAT STANDARD; PRT; 845 AA.
 ID NFM_RAT
 AC P12839; Q63370;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Neurofilament triplet M protein (160 kDa neurofilament protein)
 DE (Neurofilament medium polypeptide) (NF-M).
 GN NEF3 OR NEFM OR NFM.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OC NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87282618; PubMed=2441012;
 RA Napolitano E.W., Chin S.S.M., Colman D.R., Liem R.K.H.;
 RT "Complete amino acid sequence and in vitro expression of rat NF-M,"

RT the middle molecular weight neurofilament protein.";
 RL J. Neurosci. 7:2590-2599 (1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Wistar;
 RX MEDLINE=92332596; PubMed=1321159;
 RA Kelly B.M., Gillespie C.S., Sherman D.L., Brophy P.J.;
 RT "Schwann cells of the myelin-forming phenotype express neurofilament
 RT protein NF-M.";
 RL J. Cell Biol. 118:397-410 (1992).
 RN [3]
 RP PHOSPHORYLATION SITES, AND REVISION TO 500.
 RX MEDLINE=92165797; PubMed=1537832;
 RA Xu Z.-S., Liu W.-S., Willard M.B.;
 RT "Identification of six phosphorylation sites in the COOH-terminal
 RT tail region of the rat neurofilament protein M.";
 RL J. Biol. Chem. 267:4467-4471 (1992).
 RN [4]
 RP CARBOHYDRATE-LINKAGE SITES.
 RX MEDLINE=93346421; PubMed=8344946;
 RA Dong D.L.-Y., Xu Z.-S., Chevrier M.R., Cotter R.J., Cleveland D.W.,
 RA Hart G.W.;
 RT "Glycosylation of mammalian neurofilaments. Localization of multiple
 RT O-linked N-acetylglucosamine moieties on neurofilament polypeptides
 RT L and M.";
 RL J. Biol. Chem. 268:16679-16687 (1993).
 CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFM IS
 CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
 CC THOUGHT THAT PHOSPHORYLATION OF NFM RESULTS IN THE FORMATION OF
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
 CC OF AXONAL CALIBER.
 CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
 CC INCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
 CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M18628; AAA41696.1; -
 CC EMBL; Z12152; CAA78136.1; -
 CC PIR; A45669; A45669.
 CC Glycosylated; P12839; -
 CC InterPro; IPR006821; Filament_head.
 CC InterPro; IPR001664; IF.
 CC InterPro; IPR002957; Keratin_I.
 CC Pfam; PF00438; filament; 1.
 CC Pfam; PF04732; filament_head; 1.
 CC PRINTS; PR01248; TYPEIKERATIN.
 CC PROSITE; PS00226; IF; 1.
 CC Intermediate filament; Coiled coil; Neutrone; Phosphorylation;
 CC Glycoprotein.
 FT INIT MET 0 0
 FT DOMAIN 1 103
 FT DOMAIN 103 410
 FT DOMAIN 411 844
 FT DOMAIN 844 1033
 FT DOMAIN 1033 134
 FT DOMAIN 135 147
 FT DOMAIN 148 246
 FT DOMAIN 247 263
 FT DOMAIN 264 285
 FT DOMAIN 286 289
 FT DOMAIN 290 410
 FT CARBOHYD 47 47
 BY SIMILARITY.
 HEAD.
 TAIL.
 COIL 1A.
 LINKER 1.
 COIL 1B.
 LINKER 12.
 COIL 2A.
 LINKER 2.
 COIL 2B.
 O-LINKED (GLCNAC).
 /FTId=CAR_000130.

FT	CAREOHYD	430	430	O-LINKED (GLCNAC).
FT	MOD_RES	502	502	/FTIG-CAR 000131.
FT	MOD_RES	506	506	PHOSPHORYLATION.
FT	MOD_RES	536	536	PHOSPHORYLATION.
FT	MOD_RES	603	603	PHOSPHORYLATION.
FT	MOD_RES	608	608	PHOSPHORYLATION.
FT	MOD_RES	666	666	PHOSPHORYLATION.
FT	CONFLICT	17	17	MISSING (IN REF. 2).
FT	CONFLICT	21	21	R -> P (IN REF. 2).
FT	CONFLICT	204	204	V -> L (IN REF. 2).
FT	CONFLICT	500	500	MISSING (IN REF. 1).
FT	SEQUENCE	845 AA;	95660 MW;	316C41655B1197D CRC64;

Query Match 18.7%; Score 166.5; DB 1; Length 845;
 Best Local Similarity 30.9%; Pred. No. 0.11;
 Matches 60; Conservative 29; Mismatches 80; Indels 25; Gaps 8;

QY	6	EEAEAEAK--RRADAKGKPKGKAGVPGELATPDKENDAKSSDSVGEETLPSPS	62
DB	559	EEGETEABEGEGEAEAEKKEKTEGKVEEMAIKEIKVEKAKSPVKSPVEEVKPKPE	618
QY	63	LKPEKKVAEAEKKVEAEKKAE-----DQKEEDRRNYPTNTYKTLLEIAESDVE--	112
DB	619	AKAGKDEQKEKEKEKEKVEAKESPREKEVEKEKEKPKDVK--KKAESPVEKAVEEM	676
QY	113	---VKAELELVKEAKE-PRNEEKVKAKAVEESKKAATRLKIKITDRKAE---EE	164
DB	677	ITITSVKVSLKDKKEKPKQOEKVE-KAEEGGSEBEVG-DKSPQSKKEDIAINGE	734
QY	165	AKRAAEEDKVKEX	178
DB	735	VEGKEEEEOETQEK	748

Search completed: November 21, 2003, 13:32:33
 Job time : 6.20984 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:32:37 ; Search time 19.9667 Seconds
(without alignments) 1285.946 Million cell updates/sec

Title: US-09-298-523C-13

Perfect score: 2537
Sequence: 1 MFASKSRKRVHYSTRKFSVG.....NRLTQQDPKPKKPAQSTP 511

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 261868 seqs, 50246823 residues

Total number of hits satisfying chosen parameters: 261868

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents AA New:*
1: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep:*
2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2516	99.2	693	1	PCT-US03-27401-468 Sequence 468, App
2	2516	99.2	693	6	US-10-473-928-4598 Sequence 4598, App
3	397.5	15.7	744	1	PCT-US03-27401-251 Sequence 251, App
4	397.5	15.7	744	6	US-10-473-928-32 Sequence 32, App
5	361.5	14.2	1033	6	US-10-258-899A-3455 Sequence 3455, App
6	360	14.2	1026	6	US-10-258-899A-1487 Sequence 1487, App
7	283	11.2	916	7	US-60-490-890-475 Sequence 475, App
8	270	10.6	487	1	PCT-US03-11802-21 Sequence 21, App
9	265.5	10.5	793	7	US-60-490-890-2605 Sequence 2605, App
10	262	10.3	612	7	US-60-500-337-2473 Sequence 2473, App
11	262	10.3	612	7	US-60-500-337-2475 Sequence 2475, App
12	259.5	10.2	8625	6	US-10-679-063-21234 Sequence 21234, App
13	257.5	10.1	4684	1	PCT-US03-14382-432 Sequence 432, App
14	257.5	10.1	4684	1	PCT-US03-14382-434 Sequence 434, App
15	257.5	10.1	4684	1	PCT-US03-14382A-432 Sequence 432, App
16	257.5	10.1	4684	1	PCT-US03-14382A-434 Sequence 434, App
17	257.5	10.1	4684	7	US-60-479-073-497 Sequence 497, App
18	257.5	10.1	4684	7	US-60-479-073-506 Sequence 506, App
19	253.5	10.0	2524	7	US-60-502-656-219 Sequence 219, App
20	253.5	10.0	2524	7	US-60-512-690-253 Sequence 253, App
21	253.5	10.0	4576	1	PCT-US03-31476-281 Sequence 281, App
22	252.5	10.0	4576	1	US-60-479-073-495 Sequence 495, App
23	252	9.9	665	5	US-09-820-843B-107 Sequence 107, App
24	251	9.9	747	6	US-10-425-114A-67803 Sequence 67803, App
25	250.5	9.9	552	7	US-60-500-337-2474 Sequence 2474, App
26	250	9.9	1270	6	US-10-473-040-696 Sequence 696, App

27	250	9.9	2067	6	US-10-473-040-778 Sequence 778, App
28	247	9.7	662	7	US-60-500-337-2476 Sequence 2476, App
29	247	9.7	718	6	US-10-258-899A-3786 Sequence 3786, App
30	245.5	9.7	619	6	US-10-258-899A-1818 Sequence 1818, App
31	244.5	9.6	1435	6	US-10-473-551-661 Sequence 661, App
32	244.5	9.6	1021	6	US-10-473-551-1057 Sequence 1057, App
33	244	9.6	507	6	US-10-425-114A-47285 Sequence 47285, App
34	242.5	9.6	774	6	US-10-258-899A-3849 Sequence 3849, App
35	240	9.5	1533	6	US-10-679-063-23311 Sequence 23311, App
36	234.5	9.2	680	6	US-10-425-114A-68086 Sequence 68086, App
37	232	9.1	8943	6	US-10-679-063-20888 Sequence 20888, App
38	231	9.1	748	6	US-10-425-114A-72422 Sequence 72422, App
39	229.5	9.0	1848	6	US-10-687-046-6 Sequence 6, App
40	228.5	8.9	743	1	PCT-US02-18638A-188 Sequence 188, App
41	225	8.9	1404	6	US-10-473-576-2 Sequence 2, App
42	224.5	8.8	843	6	US-10-258-899A-1144 Sequence 1144, App
43	224.5	8.8	1879	6	US-10-296-115-1265 Sequence 1265, App
44	224.5	8.8	1960	6	US-10-258-899A-1516 Sequence 1516, App
45	224.5	8.8	1963	6	US-10-258-899A-3484 Sequence 3484, App

ALIGNMENTS

```
RESULT 1
PCT-US03-27401-468
Sequence 468, Application PC/TUS0327401
GENERAL INFORMATION:
APPLICANT: TUFTS UNIVERSITY
TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSIS, TREATMENT AND
TITLE OF INVENTION: PREVENTION OF ACTIVE INFECTION
FILE REFERENCE: 700355-52941-PCT
CURRENT APPLICATION NUMBER: PCT/US03/27401
CURRENT FILING DATE: 2003-09-02
PRIOR APPLICATION NUMBER: US 60/407,082
PRIOR FILING DATE: 2002-08-30
NUMBER OF SEQ ID NOS: 560
SOFTWARE: PatentIn version 3.2
SEQ ID NO 468
LENGTH: 693
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
PCT-US03-27401-468

Query Match 99.2%; Score 2516; DB 1; Length 693;
Best Local Similarity 99.6%; Pred. No. 1.8e-75;
Matches 511; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 1 MFASKSRKRVHYSTRKFSVG-ASVYVASLVMGSVYVATENEGATQVPTSSNRANESQAO 59
DB 1 MFASKSRKRVHYSTRKFSVGASVYVASLVMGSVYVATENEGATQVPTSSNRANESQAO 60
QY 60 GEOPKLDSDRDARKEVEYVKKIVGESYAKSTKSHHTTVALVNLNNIKNEYLNKIV 119
DB 61 GEOPKLDSDRDARKEVEYVKKIVGESYAKSTKSHHTTVALVNLNNIKNEYLNKIV 120
QY 120 ESTSESQLOILMMESRSKVDVAVSKEFKDSSSSSSSDSTKPEASDTAKPNKPTPEKV 179
DB 121 ESTSESQLOILMMESRSKVDVAVSKEFKDSSSSSSSDSTKPEASDTAKPNKPTPEKV 180
QY 180 AEAKKVVEA-KKAKOKEEDRRNYPITTKTLEIAESDVEVKAELELVYKANEPR 238
DB 181 AEAKKVVEA-KKAKOKEEDRRNYPITTKTLEIAESDVEVKAELELVYKANEPR 240
QY 239 DEOKIQAEAEVSKOAEATRLTKITDREAEAEAKRRADAEOKPKGRARGVGEL 298
DB 241 DEOKIQAEAEVSKOAEATRLTKITDREAEAEAKRRADAEOKPKGRARGVGEL 300
QY 299 ATPDKKENDAKSSDSSVGEEITLPSLKPPEKVAEKKVYEAKKKAEDQKEEDRRNYP 358
DB 301 ATPDKKENDAKSSDSSVGEEITLPSLKPPEKVAEKKVYEAKKKAEDQKEEDRRNYP 360
QY 359 NTYKTIETLEIAESDVEVKAELELVKEBAKEPNEKVKQAKAEVSKKAEATRLKIKT 418
```

|||||
Db 361 NTYKLELEIASDVVKAAELIVKEAKEPRNEKVKQAQAEVSKKAEATRLKIKT 420
Qy 419 DRKKAEEAKRKAEEEDKYKEKPAPOPAPAPKAPKAPAPKAPENPAEQPAEKPADQ 478
Db 421 DRKKAEEAKRKAEEEDKYKEKPAPOPAPAPKAPKAPAPKAPENPAEQPAEKPADQ 480
Qy 479 AEDYARSEEEYNRLTQOQPPKTEKPAQSTP 511
Db 481 AEDYARSEEEYNRLTQOQPPKTEKPAQSTP 513

RESULT 2
US-10-472-928-4598
; Sequence 4598, Application US/10472928
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926WO
; CURRENT APPLICATION NUMBER: US/10/472,928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 4598
; LENGTH: 693
; TYPE: PRN
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; OTHER INFORMATION: choline binding protein A (cbpa)
; OTHER INFORMATION: Cellular location: outside
; OTHER INFORMATION: Feature of note: WY motif
; OTHER INFORMATION: Similar to strain R6 sequence 15904036 (0..E+01)
US-10-472-928-4598

Query Match 99.2%; Score 2516; DB 6; Length 693;
Best Local Similarity 99.6%; Pred. No. 1.8e-75;
Matches 511; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Qy 1 MFASSEKRVHSIRKFSVG-ASVVVASLWGSVVAHATENEGATOPTSSNRANESQAEQ 59
Db 1 MFASSEKRVHSIRKFSVGASVVVASLWGSVVAHATENEGATOPTSSNRANESQAEQ 60
Qy 60 GEOPKLDSEDPKARKEVEEYKTVGSYAKSTKRRHTITVALVNLNNINNEYLNTKIV 119
Db 61 GEOPKLDSEDPKARKEVEEYKTVGSYAKSTKRRHTITVALVNLNNINNEYLNTKIV 120
Qy 120 ESTSSQQLIMMESRSKYDVAVSFKPKDSSSSSSSDSTKPEASDTAKPKPTPEGEKV 179
Db 121 ESTSSQQLIMMESRSKYDVAVSFKPKDSSSSSSSDSTKPEASDTAKPKPTPEGEKV 180
Qy 180 AEAKKVEEA-KKADQKEDRRNPTITTYKLELEIASDVVKAAELIVKCANER 238
Db 181 AEAKKVEEA-KKADQKEDRRNPTITTYKLELEIASDVVKAAELIVKCANER 240
Qy 239 DEOKIKQAEAVESQAEATRLKIKTDRKEAEKEERKADAKKEQKPKGRKGVPGEL 298
Db 241 DEOKIKQAEAVESQAEATRLKIKTDRKEAEKEERKADAKKEQKPKGRKGVPGEL 300
Qy 299 ATPDCKENDAKSSDSVGEETLPSPSLKEPEKVAEAEKKEVEAKKAEQKEDRRNPT 358
Db 301 ATPDCKENDAKSSDSVGEETLPSPSLKEPEKVAEAEKKEVEAKKAEQKEDRRNPT 360
Qy 359 NTYKLELEIASDVVKAAELIVKEAKEPRNEKVKQAQAEVSKKAEATRLKIKT 418
Db 361 NTYKLELEIASDVVKAAELIVKEAKEPRNEKVKQAQAEVSKKAEATRLKIKT 420
Qy 419 DRKKAEEAKRKAEEEDKYKEKPAPOPAPAPKAPKAPAPKAPENPAEQPAEKPADQ 478
Db 421 DRKKAEEAKRKAEEEDKYKEKPAPOPAPAPKAPKAPAPKAPENPAEQPAEKPADQ 480

Qy 479 AEDYARSEEEYNRLTQOQPPKTEKPAQSTP 511
Db 481 AEDYARSEEEYNRLTQOQPPKTEKPAQSTP 513

RESULT 3
PCT-US03-27401-251
; Sequence 251, Application PC/TUS0327401
; GENERAL INFORMATION:
; APPLICANT: TUFIS UNIVERSITY
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSIS, TREATMENT AND
; FILE REFERENCE: 700355-52941-PCT
; CURRENT APPLICATION NUMBER: PCT/US03/27401
; CURRENT FILING DATE: 2003-09-02
; PRIOR APPLICATION NUMBER: US 60/407,082
; PRIOR FILING DATE: 2002-08-30
; NUMBER OF SEQ ID NOS: 560
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 251
; LENGTH: 744
; TYPE: PRN
; ORGANISM: Streptococcus pneumoniae
PCT-US03-27401-251

Query Match 15.7%; Score 397.5; DB 1; Length 744;
Best Local Similarity 28.0%; Pred. No. 6.8e-07;
Matches 147; Conservative 89; Mismatches 166; Indels 123; Gaps 22;

Qy 47 TSSNRANESQAEQOEOPKLDSEDPKARKEVEEYKTVGSYAKSTKRRHTITVALVN- 105
Db 63 TAKKAEADQKKYDDQKRT-EKARKE-----AASQKNDVALVQNA 106
Qy 106 -----ELNNINNEYLNTKIVESTSSQQLIMMESRSKYDVAVSFKPKDSSSSSDSTK 160
Db 107 YKEYREVQNRSKY-----KSDAEYOKKL-----TEVDSKIEFKARKEQODLONKFEVR 155
Qy 161 PEASDTAKPKPTPEGEKVAAKKKVEAKKAKKQKEDRRNPTITTYKLELEIASDV 220
Db 156 --AVVPEPN-----ALAEKKKAEBAKA--EEKAKKRY--DYATLKVALAKKEV 200
Qy 221 EVKKAELIVKCANERDEOKIKQAEAVESQAEATRLKIKT-----KT 265
Db 201 EAKLEIEIKL-----QYEISTLEGEVATAGHQVNDLKKLAGADPDQTEVIEAKL 251
Qy 266 DREAEAEAKRKAADAKESGK-----PKGRKGV--PGLATPPKENDAKSSSS 314
Db 252 KKGAEELNAKQAEIAKKQTELEKLDLSDPBGKTQDELDKAEAEALDKADDELONKVAD 311
Qy 315 VGEETLPSPL-----KPEKVAE-----AEKKEAEAKKAEQKEDRRNPTITTYKLE 365
Db 312 LEKEISNLEILLGADPEDDTALONKLAANKAEIAKKQTELEKLDLSDPBGKTQDELD 371
Qy 366 LEIASDVEVKKAELE-LVKEAKEPRN-----EEKVKQAQAEVSKKAEATRLK 414
Db 372 KEAEAEALDKADELQNNVADLEKEISNLEILLGADSEDDTALONKLAATKAE--LE 428
Qy 415 KITDRKKAEEAEAKRKAEEEDKYKEKPAPOPAPAPKAPKAPAPKAPENPAEQPAEK 473
Db 429 KTQKELDALNALNELSPDGEER--TPAPAPQEPQAPAPAPKAPAPKAPENPAEQPAEK 486
Qy 474 PADQAEEDYARSEEEYNRLTQOQPPKTEKPAQ-----PSTP 511
Db 487 PAPAPKPEQAPAPAPPE-----QPAKPEKPAEETPOPEKPAPE 524

RESULT 4
US-10-472-928-32
; Sequence 32, Application US/10472928
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS

Query Match	15.7%	Score 397.5;	DB 6;	Length 744;
Best Local Similarity	28.0%	Pred. No. 6.8e-07;		
Matches 147;	Conservative 89;	Mismatches 166;	Indels 123;	Gaps 22

RESULT 5
US-10-258-899A-3455

```

sequence 3455, Application US/10258893A
GENERAL INFORMATION:
APPLICANT: Tang, Y. Tom
APPLICANT: Liu, Chenchua
APPLICANT: Drmanac, Redjoje T.
APPLICANT: Asundi, Vinod
APPLICANT: Zhou, Ping
APPLICANT: Xu, Chongjun
APPLICANT: Cao, Yicheng
APPLICANT: Ma, Yunqiang
APPLICANT: Zhao, Qing A.
APPLICANT: Wang, Dunruil
```

Query Match	14.2%	Score 361.5;	DB 6;	Length 1033;
Best Local Similarity	26.1%	Pred. No. 1.2e-05;		
Matches 155;	Conservative 82;	Mismatches 234;	Indels 123;	Gaps 24;

```

Oy      5 KSEKXVHYSIRKXSVGASVVV-----ASLWMSGVHATHTEMGATQVPTSSNBRANESQAQ 59
Db      445 KSEBKTI--KXVESKSEKTVIWEBOTEBTQVTEETEEEDNEKAEBCGEGGEEBAEG 502
Oy      60 GEOPKRLDSEBRDKA--RKEVEBYVK--KIYVESYAKSTPKRHHTITVALVNEIANNIKNEULN 116
Db      503 GEBETTSPPABEAAAPPEKAKSPVYKEAKSPPAEKSPKEKEAKSPAPVYKSPPEKAKSP-- 559
Oy      117 KIYESTSESQOLQILIMNESRSKVDAAVSKPEKDSSSSSSSSDSSTKPEAKSDTAKPNKP--- 172
Db      560 --AKEAKSP-----PEAKSPKDGQKONFOAYVYKSPPEKAKSPAKEAKSPAEAKSPPEKAK 612
Oy      173 ----TEPEKAEKAEKKKY--EEAKKAKQOKEEDRRNRYPIITYKTTLELTAEEDVYKAELE 228
Db      613 SPVKEEAKSPAEAKSPVYKEAKSPAPEKVPKSPPEKAKSPT--KEAKSPPEKAKSPPEKAKSP-- 669
Oy      229 LVYKKAPEPPDEOKTIKQAEBAVESK-----QAEATRLKTKITD--REAEAEAKBRADAKE 282
Db      670 --KEEAKSPPEKAKSPVYKAEKSPPEKAKSPVYKAEAKSPPEKAKSPVYKEAKSPPEKAKSPVKE 727
Oy      283 QGPKRGKAGKGVGELATPPDKKENDAKSDSDSVGEETLPSPSLKPEKKVAAEAKKYVEAK 342
Db      728 EAKSPPEKAKSPVYKEAKTPEKAKSPVYKEAKS-----PEKAKSP--KAKTLDVKSPEAK 780
Oy      343 KKAEDQCKEEDRRNRYPIITYKTTLELT-----ASSDV--EYKAELEL--VKEEAKERPNBE 394
Db      781 TPKEEAKSPADPAPPEKAKSPVYKEAKSPPEKAKSPPEKAKPAEKIKPKKEEYVYKSPVKEE 840

```

QY 5 KSERKVHYSIRKFSVGASVV-----ASLVMGSVHATENEGATQVPTSSNRANESQAEQ 59

```

QY      53  NESQAEQGEOPKPLDSEDDPKARK-----EVEEYAKIIGSEAYAST-----KKR  96
      368  NELRQTKEMARHLREYQDLLNWKVALDIEIAAYAKKLEEGEETRPSTFAGSITGPLYTHR  427
QY      97  HTIYVALYNELN-----NIKIELANKIVESTSSQQLIMESKSKYDAVSKREKKS  149
      428  PPIITSIKIQTQVAPKLVQHKRVEEIIETTKVDEKSENEEALTAITEELIASKME  487
Db

```

Qy 150 SSSSSDSTYKEAD---TAK-ENKTEPEGEVAAKKEVAAKAKDOKEEDRNYP 204
Db 488 KKEAAEKEEPEEAEVEEAAKSPVAKATAPVKEEKEEKEEBOGEEDBEGANS- 545
Qy 205 TTTTLELELAESVVEVKALELVKVKANPRDEQIKOAEAVESKOAEATRLKIK 264
Db 546 -----DOAEBOGSEKESSSE-----KEEBOGEGETEAEAEBAEAEKEEKEVE 589
Qy 265 TDREAEAEARPADAEQKRGKRAKRGVDELATPDKKENDAKSSSSVGEETLPSPS 324
Db 590 EKSEEVATKEELVADAKVE-KPE-KAKSPVPKSPVEEKSGSPVPKSPVEEKSGSPVPSP 647
Qy 325 LKPEKVAEAEKVEE-AKKAEDOKEDERNRYPTNTYKLELAESDVVEVKALELV 383
Db 648 VEEKSGSPVPKSPVEEKSGSPVSKSPVEEKAKSP-----VPSVPEEAKSKAEVG 697
Qy 384 KEAEKPENEEKVKOAKAEVESKKAATRLKIK--TURKKAEEBAKRAA----- 432
Db 698 KGEQKE-EEEKVEKPAKPEKEVEKKE-----EKPKOVPEKKAESPVEKEAVAEVVTITKS 752
Qy 433 -----EEDKYKEKPAEPOPAPAPAKAPAPKPEPAPAPAKAPADQAEEDVARR 486
Db 753 VKVHLEKTKKEEGPLQO-----EKEKEKAGEGSGSESGSDKAGKSGRREDAVN 803
Qy 487 SEEEYNRLTQOQ 498
Db 804 GEVEGKEVEQE 815

RESULT 8
PCT-US03-11802-21
; Sequence 21, Application PC/TUS0311802
; GENERAL INFORMATION:
; APPLICANT: Washington University
; TITLE OF INVENTION: Regulated Attenuation of Live Vaccines to Enhance Cross-Protective
; TITLE OF INVENTION: Immunogenicity
; FILE REFERENCE: 56029-40434
; CURRENT APPLICATION NUMBER: PCT/US03/11802
; PRIOR FILING DATE: 2003-04-15
; PRIOR APPLICATION NUMBER: US 60/373,626
; PRIOR FILING DATE: 2002-04-18
; PRIOR APPLICATION NUMBER: US 60/372,616
; PRIOR FILING DATE: 2002-04-15
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 21
; LENGTH: 487
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
PCT-US03-11802-21

Query Match 10.6%; Score 270; DB 1; Length 487;
Best Local Similarity 28.6%; Pred. No. 0.0069;
Matches 127; Conservative 66; Mismatches 165; Indels 86; Gaps 23;
Qy 119 VESTSESOLOLIMMESKVDKAVS-KFEKSSSSSSDSTKPEASDTAKPNKPTERGE 177
Db 1 LQANESQK---EADKTKKATQKDEAAAFATITRTTIVPEPSBELATKKKAEAT 56
Qy 178 KVAE-AKKVVEA-----KKAQKEDERNRYPTTYKLEL-----EIASD 219
Db 57 KEAVAKKSEAEAEVEKKNKILLEDANEKIDIVLQNKVALLKGIAPYQNVAEELN 116
Qy 220 VEVKAELELVKVANPRD-----EOKIKOAEAVESKQAEATRLKIKITDREAE 273
Db 117 KEIARLOSIDLDAENNVEDIYKGLQAITNKKAELATQ---QNDIKQJQJLEDALE 173
Qy 274 AKR-RDAKGEKRGKRA-KRGVPEL-----ATPKKENDAKSSDSSVGEETLP 322
Db 174 LEKVLATLDPGKQJQJDELEKAEAELEKVEALQNOVALEELSLKLELNLKDAEFLQS 233
Qy 323 P---SLPEKVAEAEKVEBAKKAED-OKEED-----RNNYPTNTYKLEL---LEIAE 370

Db 234 PVASQKAEKDYDAKADKAKNAKKAVEDAQALDDAKAQQKDYEDQKTEEKALEKRA 293
Qy 371 SDVVEVKALELVVEE-AKEPRNEEKVKOAKAEV-ESKKAATRLKIKITDR----- 420
Db 294 SE-EMDYA-VAVQOAVIATQOATDKAKADADKXIDAEKAREBAKTFTNTVAMVPE 351
Qy 421 -----KKAEEBAKRA-----AEEDKYKEKPAEPOPAPAPK-AEKPAAPK---P 462
Db 352 PEQALFKKXSEAEKQKAPETITKLEBAKALDEBAEKATTAQKXVDAEVAPOAKIAEL 411
Qy 463 ENPAEQKAE-KPADQOAEEDYAR 485
Db 412 ENQVHLEQELKEIDESESEEDYAK 435

RESULT 9
US-60-490-2605
; Sequence 2605, Application US/60490890
; GENERAL INFORMATION:
; APPLICANT: Li, Martha
; APPLICANT: Rupnow, Brent A.
; APPLICANT: Webster, Kevin R.
; APPLICANT: Jackson, Donald
; APPLICANT: Wong, Tai W.
; TITLE OF INVENTION: BIOMARKERS OF CYCLIN-DEPENDENT KINASE MODULATION
; FILE REFERENCE: D0310 PSP
; CURRENT APPLICATION NUMBER: US/60/490,890
; PRIOR FILING DATE: 2003-07-29
; NUMBER OF SEQ ID NOS: 2779
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2605
; LENGTH: 793
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-490-890-2605

Query Match 10.5%; Score 265.5; DB 7; Length 793;
Best Local Similarity 22.2%; Pred. No. 0.013;
Matches 123; Conservative 105; Mismatches 192; Indels 135; Gaps 23;
Qy 53 NESQAEQEQPKLDSERDKARKEVEEYKVIAGESYAKSTKKHTTTVALVNLNINIK 112
Db 30 NDDDEBAERERRRARRQERLQKQEE-----ESLGQVTDQ-----VEVNAQNSVPD 76
Qy 113 E-----YLNKIVESTSSQJ-QILMESR-SKVDAVSKFEK-----DSSS----- 152
Db 77 EBAKTTTNTQVGGDDAEALERLARREERRQKRLQELERQKDFPTITDASISLPSRR 136
Qy 153 -----SSSDSTKPEASDTAKPNKPTERGEVAAKKE-----VEEAKKAKDOKEEDRN 202
Db 137 MQNDTAENETTEREKESESQERYELETETVTKSYQKNDWRDAEENKKEDKEKEE 196
Qy 203 YP---TTTYTLELEI-----AESDVVEVKALELVKVANPRDEQIKOAEAEV----- 250
Db 197 KPKRGSIGENQVVMVESEKTTESQEBTVVMSLKNQJSSSEPEKOEEREGSDLSHHEK 256
Qy 251 -----ESKQAEATRLKIKITDREBAEBAK-----RRADAKQKPKKRAKGVPEBLA 299
Db 257 MEEDKEKRAAEERARAEAEERERIKAEQDKIADERARIAEAEKAAQOERERREAE-ERE 315
Qy 300 TPDKENDAKSSDSSVGEETLPSPSLKPEKVAEAEKVEAKKAEDQ-----KEEDRN 355
Db 316 RMRBEERAAEERORIEE-----EKRAAEERRIEERERAAEERORIEEERKRA 366
Qy 356 YPTNTYKLELELAESDVVEVKALELVKEAEKPEPRNEEKVKOAKAE-----VESKKA 408
Db 367 AEERQARAE-EEEKKAEVEQKRNKQJ--EKKRAWQETIKIKEEKVEQKIEGKVVNKKKA 423
Qy 409 EATRLK---KITDRKKAEEBAKRAAEEDK-----VK-EKPAEPOPAPAPKA--- 453
Db 424 QEDRLQTAVALKKQGEKGTQVQAKREKLQEDKPTFKKEELKDKKIKDKPEKVEKFSMD 483
Qy 454 -----EKPAAPKPEENPAEQKAEKPAQOAEEDVARR 486

Db 484 RKGFTVKSQNGEFMTHTKHTENTFSRPGSRASVDPTKEAGAPQVABGRLEBRRR 543
 Qy 487 SE---EYNNRLTQOO 498
 Db 544 GETESEFEKLOKQ 558

RESULT 10

US-60-500-337-2473
 ; Sequence 2473, Application US/60500337
 ; GENERAL INFORMATION:
 ; APPLICANT: CARGILL, Michele
 ; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
 ; TITLE OF INVENTION: ENCODING HUMAN KINASE PROTEINS, METHODS OF DETECTION AND
 ; TITLE OF INVENTION: USES THEREOF
 ; FILE REFERENCE: CLO01483
 ; CURRENT APPLICATION NUMBER: US/60/500.337
 ; CURRENT FILING DATE: 2003-09-05
 ; NUMBER OF SEQ ID NOS: 123188
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2473
 ; LENGTH: 612
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-60-500-337-2473

Query Match 10.3%; Score 262; DB 7; Length 612;
 Best Local Similarity 23.9%; Pred. No. 0.014;
 Matches 124; Conservative 80; Mismatches 197; Indels 118; Gaps 19;

Qy 55 SOAEGEQPKLDSERDKARKEVEYVKIYESYAKSTKKRHTTTVALVNLNNIKREY 114
 Db 5 SOKAEIKEMLASDDEEDVSSKVEKAYVPLTG---TVKGRFA-----EMEKOROE 52
 Qy 115 LNKIVESTESQLOILMESRSKVDEAVSKFEKSSSSSSSTKPEASD---TAKPN 170
 Db 53 QKRTTEERKRRIQDMLERKRIQRELAQRAQIEDINTGTESASEGDDSLITVVPV 112
 Qy 171 KPTPEGEVAAKAKVVEAKAKADQK---EEDRRNYPTTTY-----KTLELEI 215
 Db 113 KSYTSGM---KNFEDLEKERERIKYEEDKR---IRYEQRPSLKEAKCLSLVM 165
 Qy 216 -AESDVEYKKALELVKVK-----ANPRDEQIKQAEAV-----ESKQ-BAATRLKI 263
 Db 166 DDEIESEKKSLSLPGKULTFEELERQRENKQAEERKALEEKRAFEEARRQMV 225
 Qy 264 KTDREAE-----EAKRRADAKEQCKPKGRKGVPGE-LATPDKE 305
 Db 226 NEDENQDTAKIFKGYRPGKULTSFEEMERQREDEKRAEEARRRIEEKKAFABARR 285
 Qy 306 NDAKSSDSSVGEETLPSPSLKPEKVAEAE---KVEEAKKADQKEEDRRNYPTNTY 361
 Db 286 NMVVDSDSPENYKITSQEFLLPGKLEINFELLKQMEEEKRREERKH----- 335
 Qy 362 KTLLEIAESDVEYKKALELVKEAKEPRN-----EEVKQAKAEVSKKAAEA 410
 Db 336 -----KLEMEKQFEQOLRQEMGEEBEENETFGLSREYBELIKTKRSGSIQAKLK 385
 Qy 411 TRLEKI-----KTDKKAEEBAKRAAEEDKVEKPAE-----QPPAPAPKAEKAPA 459
 Db 386 SKFEKIGLSEKEIOKRIEERARRAIDLEIKEREANFNHEEDVDVPRPARKSE---A 441
 Qy 460 PKPENPAEQPAKEKPADQQAEDYARSEEEYNRLTQOO 498
 Db 442 PFTHKVMKARFEQMAKAREEEOQRRIEQKLLRMQFEQ 480

RESULT 11
 US-60-500-337-2475
 ; Sequence 2475, Application US/60500337
 ; GENERAL INFORMATION:
 ; APPLICANT: CARGILL, Michele

; TITLE OF INVENTION: POLYMORPHISMS IN NUCLEIC ACID MOLECULES
 ; TITLE OF INVENTION: ENCODING HUMAN KINASE PROTEINS, METHODS OF DETECTION AND
 ; TITLE OF INVENTION: USES THEREOF
 ; FILE REFERENCE: CLO01483
 ; CURRENT APPLICATION NUMBER: US/60/500.337
 ; CURRENT FILING DATE: 2003-09-05
 ; NUMBER OF SEQ ID NOS: 123188
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 2475
 ; LENGTH: 612
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-60-500-337-2475

Query Match 10.3%; Score 262; DB 7; Length 612;
 Best Local Similarity 23.9%; Pred. No. 0.014;
 Matches 124; Conservative 80; Mismatches 197; Indels 118; Gaps 19;

Qy 55 SOAEGEQPKLDSERDKARKEVEYVKIYESYAKSTKKRHTTTVALVNLNNIKREY 114
 Db 5 SOKAEIKEMLASDDEEDVSSKVEKAYVPLTG---TVKGRFA-----EMEKOROE 52
 Qy 115 LNKIVESTESQLOILMESRSKVDEAVSKFEKSSSSSSSTKPEASD---TAKPN 170
 Db 53 QKRTTEERKRRIQDMLERKRIQRELAQRAQIEDINTGTESASEGDDSLITVVPV 112
 Qy 171 KPTPEGEVAAKAKVVEAKAKADQK---EEDRRNYPTTTY-----KTLELEI 215
 Db 113 KSYTSGM---KNFEDLEKERERIKYEEDKR---IRYEQRPSLKEAKCLSLVM 165
 Qy 216 -AESDVEYKKALELVKVK-----ANPRDEQIKQAEAV-----ESKQ-BAATRLKI 263
 Db 166 DDEIESEKKSLSLPGKULTFEELERQRENKQAEERKALEEKRAFEEARRQMV 225
 Qy 264 KTDREAE-----EAKRRADAKEQCKPKGRKGVPGE-LATPDKE 305
 Db 226 NEDENQDTAKIFKGYRPGKULTSFEEMERQREDEKRAEEARRRIEEKKAFABARR 285
 Qy 306 NDAKSSDSSVGEETLPSPSLKPEKVAEAE---KVEEAKKADQKEEDRRNYPTNTY 361
 Db 286 NMVVDSDSPENYKITSQEFLLPGKLEINFELLKQMEEEKRREERKH----- 335
 Qy 362 KTLLEIAESDVEYKKALELVKEAKEPRN-----EEVKQAKAEVSKKAAEA 410
 Db 336 -----KLEMEKQFEQOLRQEMGEEBEENETFGLSREYBELIKTKRSGSIQAKLK 385
 Qy 411 TRLEKI-----KTDKKAEEBAKRAAEEDKVEKPAE-----QPPAPAPKAEKAPA 459
 Db 386 SKFEKIGLSEKEIOKRIEERARRAIDLEIKEREANFNHEEDVDVPRPARKSE---A 441
 Qy 460 PKPENPAEQPAKEKPADQQAEDYARSEEEYNRLTQOO 498
 Db 442 PFTHKVMKARFEQMAKAREEEOQRRIEQKLLRMQFEQ 480

RESULT 12
 US-10-679-063-21234
 ; Sequence 21234, Application US/10679063
 ; GENERAL INFORMATION:
 ; APPLICANT: Edgerton, Michael D
 ; TITLE OF INVENTION: TRANSGENIC PLANTS WITH IMPROVED PHENOTYPES
 ; FILE REFERENCE: 38-15(52054)B
 ; CURRENT APPLICATION NUMBER: US/10/679.063
 ; CURRENT FILING DATE: 2003-10-02
 ; PRIOR APPLICATION NUMBER: 60/415,758
 ; PRIOR FILING DATE: 2002-10-02
 ; NUMBER OF SEQ ID NOS: 27373
 ; SEQ ID NO 21234
 ; LENGTH: 8625
 ; TYPE: PRT
 ; ORGANISM: Procamburus clarkii
 US-10-679-063-21234

Query Match 10.2%; Score 259.5; DB 6; Length 8625;
 Best Local Similarity 22.9%; Pred. No. 0.092;
 Matches 119; Conservative 84; Mismatches 200; Indels 117; Gaps 17;

QY 77 VEEVYKIVGESYAKSTK-----RHITVALVNLNNIKYLNKIVE 120
 DB 659 VSVFNRKVICVISTSKPACSWFKDSSKVLDRHCV-----NINEBGGKV- 708
 QY 121 STSGLDILMMESKVD-----EAVSKFEKDSSSSSDSTKREASDTAKPKPTPG 176
 DB 709 ---VOMMLKAE---KWDGSGYKVAALNKGKETTFFARGAHHRMLEKEKEKK- 756
 QY 177 EKVAEAKKVEEAKKADQKEEDRRNYPTITYKLEIAESDVEYKAELELVKANE 236
 DB 757 EEVEEKEKQEQEAAQVEEESKKEEAVKQDKPKREIKKEAPVDTFTYOLKVKPKKEE 816
 QY 237 PRDE-QIKQAEAVESKQA---EATRLKKIKTDREAEAEAKRADAKEQKPRGRK 291
 DB 817 AADBAQOQIELKIPDKQOAVLEGEKVKLKVTAKEEILTIEAQVKVQLOKVGAKEQQAVE 876
 QY 292 RGVGELATPDKKENDAKSSDSVGEETLPPSPSLKPEKVAEAEK-KVEEAKKQAEQKE 350
 DB 877 EGEVYKLVKRAKEQATEEAQV---OLKRVGAKEQOAVEEGEKVKLKVRAR-EEQAT 932
 QY 351 EDRRNYPTNTYKLEIAESDVEYKAELELVKEEAKEPNNEEKVKQ--AKAEVESKKA 408
 DB 933 EEAQVQKVKVGAEEQOAVEGEKVKLKVGAKKEQAVEBERAKLVKVGAKKEAVEEA 992
 QY 409 EATRLKIKTDKKAEEBAKRAAEEDKVKERPAEQOPAPAPRAE-----454
 DB 993 EKVOLKKEIVKVAKEEAEKETMESLQKPKIVE-PRSDAPKIEVLIKDSPPRRGSLA 1051
 QY 455 -----KPAAPKPEPNPA-----EQPKAE-----472
 DB 1052 TGSPPSRGSLVPPPEEGAGRRPSLLIADDEQKTRPGEBAVEFKGKRLRPGEEIVG 1111
 QY 473 ----KPADQQAEEYARSEEEYARLQOQPPKTEKPAQ 508
 DB 1112 KGKLVKGEQATKQ--RRRPSDVRPPSVQDDERKMDKCTP 1149

RESULT 13
 PCT-US03-14382-432
 ; Sequence 432, Application PC/TUS0314382
 ; GENERAL INFORMATION:
 ; APPLICANT: Emory University
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING
 ; FILE REFERENCE: 12804-011W01
 ; CURRENT APPLICATION NUMBER: PCT/US03/14382
 ; CURRENT FILING DATE: 2003-05-07
 ; PRIOR FILING DATE: 2002-05-07
 ; NUMBER OF SEQ ID NOS: 501
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 432
 ; LENGTH: 4684
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 PCT-US03-14382-432

Query Match 10.1%; Score 257.5; DB 1; Length 4684;
 Best Local Similarity 20.6%; Pred. No. 0.073;
 Matches 117; Conservative 125; Mismatches 196; Indels 129; Gaps 21;

QY 40 EGATQVPTSSNRANES-----QAEQGEOPKKT-DSEEDKARKEVEEYKIVGESYAKST 93
 DB 1763 EAERQLERMQLKANEAIRLQLAEVILQOKSLAQAEAKKEKEAEAREAR-RGKAEQAV 1821
 QY 94 KKGHTTVALVNLNNIKYLNKIVESTSESQ-----QILMM-----ESRSKVDEA 141
 DB 1822 RQR-----ELAQELEKORQLAEGTAQORLAAGELIRLRAETEGEQORQLLEE 1872

QY 142 VSKFEKDSSSSS-----SSDSTKREASDTAKPKNP-----172
 DB 1873 LARLQREAAATQYKQOELEAEALAVRAEMEVILASKAEESNSTSEKSKORLEBAEGR 1932
 QY 173 -TEDEKVAEAKKVEEAKKADQKEEDRRNYPTITYKLEIA-----ESDVE 221
 DB 1933 FRELAEEARLRALAEAKRQRLAEDDAQRAEAEVILAKLAIAGEATRLKTEAIA 1992
 QY 222 VKKAELELVKVA-----NEPDEQIKQAEAVESKQAEATRLKKIKTDREE--AEE 272
 DB 1993 LKEKAEAEERLRLAEDEAFQRRRLFEQAAQHKADIERLRLQLKASPSLEERQGLVED 2052
 QY 273 EAKRADAKEO-----GKPK-----GRARGVGEATPDKKENDAKSSSS 314
 DB 2053 TLRRQVEEETLALKASFEEAAGKAELELELIRISNAEDTLRSKEQAELEAARQRL 2112
 QY 315 VGEETLPPSLKPEKVAEAEKVEEAKKADQKEEDRRNYPTNTYKLEIAESDVE 374
 DB 2113 AAE-----ERRRREAEERVO--KSLAEAEARQKALAEVVERLAKYBEARSL 2161
 QY 375 VKKAE-----LELVKEEA-KEPRNEK-----VKQAEVESKKAATRLKIKTDREK 422
 DB 2162 RERAEQESARQLQIAQEAQKRLQAEKKAFAVQOKEOEIQ-----QTLQEOGVLDRLR 2217
 QY 423 AEEBAKRAAEEDKVKERPAEQP-----OPAPKAEKPAKPEPNPAEQPAEKPAD 476
 DB 2218 GEAEARRAAEAEAEAVQAEAREASRRQVEEAEERLKQSHQAEQAPAAQAAAEK-LR 2276
 QY 477 QQAEEYARSEEEYARLQOQPPKTE 503
 DB 2277 KEAQEAEARRAQAEQALRQKQAAAE 2303

RESULT 14
 PCT-US03-14382-434
 ; Sequence 434, Application PC/TUS0314382
 ; GENERAL INFORMATION:
 ; APPLICANT: Emory University
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING
 ; FILE REFERENCE: 12804-011W01
 ; CURRENT APPLICATION NUMBER: PCT/US03/14382
 ; CURRENT FILING DATE: 2003-05-07
 ; PRIOR FILING DATE: 2002-05-07
 ; NUMBER OF SEQ ID NOS: 501
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 434
 ; LENGTH: 4684
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 PCT-US03-14382-434

Query Match 10.1%; Score 257.5; DB 1; Length 4684;
 Best Local Similarity 20.6%; Pred. No. 0.073;
 Matches 117; Conservative 125; Mismatches 196; Indels 129; Gaps 21;

QY 40 EGATQVPTSSNRANES-----QAEQGEOPKKT-DSEEDKARKEVEEYKIVGESYAKST 93
 DB 1763 EAERQLERMQLKANEAIRLQLAEVILQOKSLAQAEAKKEKEAEAREAR-RGKAEQAV 1821
 QY 94 KKGHTTVALVNLNNIKYLNKIVESTSESQ-----QILMM-----ESRSKVDEA 141
 DB 1822 RQR-----ELAQELEKORQLAEGTAQORLAAGELIRLRAETEGEQORQLLEE 1872
 QY 142 VSKFEKDSSSSS-----SSDSTKREASDTAKPKNP-----172
 DB 1873 LARLQREAAATQYKQOELEAEALAVRAEMEVILASKAEESNSTSEKSKORLEBAEGR 1932
 QY 173 -TEDEKVAEAKKVEEAKKADQKEEDRRNYPTITYKLEIA-----ESDVE 221
 DB 1933 FRELAEEARLRALAEAKRQRLAEDDAQRAEAEVILAKLAIAGEATRLKTEAIA 1992
 QY 222 VKKAELELVKVA-----NEPDEQIKQAEAVESKQAEATRLKKIKTDREE--AEE 272
 DB 1993 LKEKAEAEERLRLAEDEAFQRRRLFEQAAQHKADIERLRLQLKASPSLEERQGLVED 2052
 QY 273 EAKRADAKEO-----GKPK-----GRARGVGEATPDKKENDAKSSSS 314
 DB 2053 TLRRQVEEETLALKASFEEAAGKAELELELIRISNAEDTLRSKEQAELEAARQRL 2112
 QY 315 VGEETLPPSLKPEKVAEAEKVEEAKKADQKEEDRRNYPTNTYKLEIAESDVE 374
 DB 2113 AAE-----ERRRREAEERVO--KSLAEAEARQKALAEVVERLAKYBEARSL 2161
 QY 375 VKKAE-----LELVKEEA-KEPRNEK-----VKQAEVESKKAATRLKIKTDREK 422
 DB 2162 RERAEQESARQLQIAQEAQKRLQAEKKAFAVQOKEOEIQ-----QTLQEOGVLDRLR 2217
 QY 423 AEEBAKRAAEEDKVKERPAEQP-----OPAPKAEKPAKPEPNPAEQPAEKPAD 476
 DB 2218 GEAEARRAAEAEAEAVQAEAREASRRQVEEAEERLKQSHQAEQAPAAQAAAEK-LR 2276
 QY 477 QQAEEYARSEEEYARLQOQPPKTE 503
 DB 2277 KEAQEAEARRAQAEQALRQKQAAAE 2303

```

OY      222  VKKALTELTVKVA-----NEPPOKOTKQAEAVESQOAPATLKKITKTREE--ABE 272
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1993  LKEKAENEBRLRLIAEDFAFORRLIEBOAQHKQITIEERLALTRKASSELERGVLVED
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      273  BAKRPADAKEQ-----GKPK-----GAKRGVPBELATPDPKENDAKSSDS 314
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2053  TLRQGRQYBEEITLAKSFEPKAAAKALELELGRIRSNABEDTLASKXQALEHARQOQL 211
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      315  VGEETLPSBPLKPEKKVAEAEKKVEEAKKADQKEEDRNYPTNTYKTELEIAESDVE 374
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2113  AAEE-----ERRREAEERVO--KSLAAEEBAQRKALTEVERLTKAVEARSL 216
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      375  VKKAE-----LELYKEA--KEPRNEEK-----VKQAKAEVSKQAELTRLEKITDCK 422
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2162  RERAEQESARQIOLQOEAQOKQLQAEKHAFAVQOKEOLO-----QTLQOEOSVLDBLR 221
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      423  ABEKKAERKAAEBDKYKEKPAEDP-----QPAKPAKEPRAPAPKPNPAEDPKAKERAD 476
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2218  GEAEKARRPAAEEAEARVQAEERAAQSSRRQYAEERLTKQSAEBQAOABQAOQAAEK--LR 227
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY      477  QOAEEDVARRSSEENRILTQOOPPKTE 503
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      2277  KEAEQEAARRAQEQALRQKQAABE 2303

```

```

RESULT 15
PCT-US03-14382A-432
; Sequence 432, Application PC/TUS0314382A
; GENERAL INFORMATION:
; APPLICANT: Emory University
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING
; TITLE OF INVENTION: ANTI-VIRAL AGENTS
; FILE REFERENCE: 12804-011WO1
; CURRENT APPLICATION NUMBER: PCT/US03/14382A
; CURRENT FILING DATE: 2003-08-20
; PRIOR APPLICATION NUMBER: US 60/378,711
; PRIOR FILING DATE: 2002-05-07
; NUMBER OF SEQ ID NOS: 501
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 432
; LENGTH: 4684
; TYPE: prf
; ORGANISM: Homo sapiens
PCT-US03-14382A-432

```

Query Match	10.1%;	Score 257.5;	DB 1;	Length 4684;
Best Local Similarity	20.6%;	Pred. No. 0.073;		
Matches 117;	Conservative 125;	Mismatches 196;	Indels 129;	Gaps 21;

Qy	40	CGATGCTVPSNRANS-----QAEQGEQPKTL-DESRQKAREVEAYKKIVGEEVAST	93
Db	1763	EAEROLIEHQKANEALRLRLQAEELVLOOKSLAQAEAKQKEAEAREARR-RGAEEQAV	1822
Qy	94	KRHRTTALVNEELNNIKNEYLNKIVESTESQTL-----QILMM-----ESRSKVDEA	141
Db	1822	RQR-----ELAEQLEKQROLAEQTAQORLLAEQELRLRAETEGSQQRQLLEE	1877
Qy	142	VSKEPKDSSSS-----SSDSSTKPEASDTAKPNP-----	172
Db	1873	LARLQREAAATQQRQLELAELAKVRAEMEVLLASKAKAAEESRSTSEKSKQRLAEAGR	1932
Qy	173	TEPEGEKVAEAKKVEEAKKADQKEEDRRNYPITTYTLELEIA-----ESVE	221
Db	1933	FRELAEEEAARLRALAEAEAKRQRLAEEDPAARQRAEAERVLAEKLLAIGATRLKTEAIEA	1992
Qy	222	VKKAELELVKRYA-----NEPRQEKIKQAAEVEKQAEATRLKKIKITDRER-AEE	272
Db	1993	LKEKEAENERRLRRLAEDEAFQRRRLLEEQAQHKADIERLRLQRLVRAASDELERQGLVED	2052
Qy	273	EAKKRAADAKQ-----GKPR-----GPAKRGVPGELATPDKKNDKASDSS	314
Db	2053	TLRQPROVEEITLAKKAFEPKAAAGKALELELGTGRSNAEDTLKSKQALEAARQRL	2112

```

QY 315 VGEETLPSPSTKPEKKVAAEKKEAAKKKDDCKEDBRNNPTNTYKLTLEIIESDVE 374
Db 2113 AAEE-----ERRRRAEERVO--KSLAAEEBAQRKALBEVEVLKAKVEARSL 216
QY 375 VKAE-----LELVKEBA-KEPRNEEK-----VKARAEEVSKKAENTLEIKITDRKK 422
Db 2162 RERAEQESARQOTQLAQAAROKRLQAEERKAAHFAVQCKEQEID-----QTLQEQSVLDRUR 221
QY 423 AAEERKRAAEEDDKYKEKPAEQ-----QAPAPAKAKKPAAPKPEPNAPQKAEKPPD 476
Db 2218 GEAEEARPAEAEEAEARVQAEEREAQAOSRRQVEAEERLKQSAEEQAQAAPAAAEK-LR 227
QY 477 QOAEEDVARSSEEEYNRLTQOQPPKTE 503
Db 2277 KEAEQEAARRRQAEEQALRQKQADAE 2303

```

Search completed: November 21, 2003, 13:41:38
Job time : 22.9667 secs

Query Match	100.0%	Score 2537	DB 16	Length 511
Best Local Similarity	100.0%	Pred. No. 1.6e-129		
Matches 511	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	MFASKSRKHYSTRKPSVGASVYVSLIWMGSSVHATENEGATQVPTSSNRANESQAEQG	60	

```
Db      1 MFASKSERKRVHSIRKFSVGSVVAASLVMSGVVHATENEGATQVPTSSNRANESQAEQG 60
Qy      61 EOPKLDSEBDKARKEVEEYVKKIIGESYAKSTKRHTTTVALVNLNINIKNEYLINKIVE 120
Db      61 EOPKLDSEBDKARKEVEEYVKKIIGESYAKSTKRHTTTVALVNLNINIKNEYLINKIVE 120
Qy      121 STSESQQLIMMESRSKYDEAVSKFEKDSSTSSSDSSTKPEASDTAKPNKPTBGEKVA 180
Db      121 STSESQQLIMMESRSKYDEAVSKFEKDSSTSSSDSSTKPEASDTAKPNKPTBGEKVA 180
Qy      181 EAKKVEBEAKKAKQKEDRNNYPTTYYKTLELEIAESDVEYKKALELVKYNANEPDE 240
Db      181 EAKKVEBEAKKAKQKEDRNNYPTTYYKTLELEIAESDVEYKKALELVKYNANEPDE 240
Qy      241 OKIKQAEAEVSKQAEATRLKKIKTDREAEAEARADAKQOGPKGRAGVPGELAT 300
Db      241 OKIKQAEAEVSKQAEATRLKKIKTDREAEAEARADAKQOGPKGRAGVPGELAT 300
Qy      301 PDKKENDAKSSDSSVGEETLPSPLKPEKVAEAKKVEAKKAEADQKEDRNNYPTNT 360
Db      301 PDKKENDAKSSDSSVGEETLPSPLKPEKVAEAKKVEAKKAEADQKEDRNNYPTNT 360
Qy      361 YKTLELEIAESDVEYKKALELVKEEAKEPNEBEKVQAKAEVSKKAATRLKIKTDR 420
Db      361 YKTLELEIAESDVEYKKALELVKEEAKEPNEBEKVQAKAEVSKKAATRLKIKTDR 420
Qy      421 KKAEEAEARKAAEEDKVEKPAEQOPAPAPAEKPAAPKPNENAEQPKAEKPADQAE 480
Db      421 KKAEEAEARKAAEEDKVEKPAEQOPAPAPAEKPAAPKPNENAEQPKAEKPADQAE 480
Qy      481 EDYARSEEEYNRILTQOQPPKTEKPAQSTP 511
Db      481 EDYARSEEEYNRILTQOQPPKTEKPAQSTP 511

RESULT 2
US-09-298-523C-13
; Sequence 13, Application US/09298523C
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/298,523C
; CURRENT FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.1
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-298-523C-13
```

```
Query Match      100.0%; Score 2537; DB 16; Length 511;
Best Local Similarity 100.0%; Pred. No. 1,6e-129;
Matches 511; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 MFASKSERKRVHSIRKFSVGSVVAASLVMSGVVHATENEGATQVPTSSNRANESQAEQG 60
Db      1 MFASKSERKRVHSIRKFSVGSVVAASLVMSGVVHATENEGATQVPTSSNRANESQAEQG 60
Qy      61 EOPKLDSEBDKARKEVEEYVKKIIGESYAKSTKRHTTTVALVNLNINIKNEYLINKIVE 120
Db      61 EOPKLDSEBDKARKEVEEYVKKIIGESYAKSTKRHTTTVALVNLNINIKNEYLINKIVE 120
Qy      121 STSESQQLIMMESRSKYDEAVSKFEKDSSTSSSDSSTKPEASDTAKPNKPTBGEKVA 180
Db      121 STSESQQLIMMESRSKYDEAVSKFEKDSSTSSSDSSTKPEASDTAKPNKPTBGEKVA 180
Qy      181 EAKKVEBEAKKAKQKEDRNNYPTTYYKTLELEIAESDVEYKKALELVKYNANEPDE 240
Db      181 EAKKVEBEAKKAKQKEDRNNYPTTYYKTLELEIAESDVEYKKALELVKYNANEPDE 240
```

```
Qy      241 OKIKQAEAEVSKQAEATRLKKIKTDREAEAEARADAKQOGPKGRAGVPGELAT 300
Db      241 OKIKQAEAEVSKQAEATRLKKIKTDREAEAEARADAKQOGPKGRAGVPGELAT 300
Qy      301 PDKKENDAKSSDSSVGEETLPSPLKPEKVAEAKKVEAKKAEADQKEDRNNYPTNT 360
Db      301 PDKKENDAKSSDSSVGEETLPSPLKPEKVAEAKKVEAKKAEADQKEDRNNYPTNT 360
Qy      361 YKTLELEIAESDVEYKKALELVKEEAKEPNEBEKVQAKAEVSKKAATRLKIKTDR 420
Db      361 YKTLELEIAESDVEYKKALELVKEEAKEPNEBEKVQAKAEVSKKAATRLKIKTDR 420
Qy      421 KKAEEAEARKAAEEDKVEKPAEQOPAPAPAEKPAAPKPNENAEQPKAEKPADQAE 480
Db      421 KKAEEAEARKAAEEDKVEKPAEQOPAPAPAEKPAAPKPNENAEQPKAEKPADQAE 480
Qy      481 EDYARSEEEYNRILTQOQPPKTEKPAQSTP 511
Db      481 EDYARSEEEYNRILTQOQPPKTEKPAQSTP 511

RESULT 3
US-09-748-875-13
; Sequence 13, Application US/09748875
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/748,875
; CURRENT FILING DATE: 2000-12-26
; PRIOR APPLICATION NUMBER: 09/298,523
; PRIOR FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-748-875-13
```

```
Query Match      100.0%; Score 2537; DB 21; Length 511;
Best Local Similarity 100.0%; Pred. No. 1,6e-129;
Matches 511; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 MFASKSERKRVHSIRKFSVGSVVAASLVMSGVVHATENEGATQVPTSSNRANESQAEQG 60
Db      1 MFASKSERKRVHSIRKFSVGSVVAASLVMSGVVHATENEGATQVPTSSNRANESQAEQG 60
Qy      61 EOPKLDSEBDKARKEVEEYVKKIIGESYAKSTKRHTTTVALVNLNINIKNEYLINKIVE 120
Db      61 EOPKLDSEBDKARKEVEEYVKKIIGESYAKSTKRHTTTVALVNLNINIKNEYLINKIVE 120
Qy      121 STSESQQLIMMESRSKYDEAVSKFEKDSSTSSSDSSTKPEASDTAKPNKPTBGEKVA 180
Db      121 STSESQQLIMMESRSKYDEAVSKFEKDSSTSSSDSSTKPEASDTAKPNKPTBGEKVA 180
Qy      181 EAKKVEBEAKKAKQKEDRNNYPTTYYKTLELEIAESDVEYKKALELVKYNANEPDE 240
Db      181 EAKKVEBEAKKAKQKEDRNNYPTTYYKTLELEIAESDVEYKKALELVKYNANEPDE 240
Qy      241 OKIKQAEAEVSKQAEATRLKKIKTDREAEAEARADAKQOGPKGRAGVPGELAT 300
Db      241 OKIKQAEAEVSKQAEATRLKKIKTDREAEAEARADAKQOGPKGRAGVPGELAT 300
Qy      301 PDKKENDAKSSDSSVGEETLPSPLKPEKVAEAKKVEAKKAEADQKEDRNNYPTNT 360
Db      301 PDKKENDAKSSDSSVGEETLPSPLKPEKVAEAKKVEAKKAEADQKEDRNNYPTNT 360
Qy      361 YKTLELEIAESDVEYKKALELVKEEAKEPNEBEKVQAKAEVSKKAATRLKIKTDR 420
Db      361 YKTLELEIAESDVEYKKALELVKEEAKEPNEBEKVQAKAEVSKKAATRLKIKTDR 420
```

QY 421 KKAEEBAKKAEEBKVKEKPAEOPAPAPAKPAEKPAPAKPENPAEQKAEKPADQOAE 480
DB 421 KKAEEBAKKAEEBKVKEKPAEOPAPAPAKPAEKPAPAKPENPAEQKAEKPADQOAE 480
QY 481 EDYARSEEEYNRLTQOOPKTEKPAQPSSTP 511
DB 481 EDYARSEEEYNRLTQOOPKTEKPAQPSSTP 511

RESULT 4

US-10-341-201-13
; Sequence 13, Application US/10341201
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; TITLE OF INVENTION: AND STRAINS THEREOF AND USES THEREFOR
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/10/341,201
; CURRENT FILING DATE: 2003-01-13
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-341-201-13

Query Match 100.0%; Score 2537; DB 29; Length 511;
Best Local Similarity 100.0%; Pred. No. 1,6e-129;
Matches 511; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPASSEKRVHYSTIRKFSVGASVVAASLVMSGVHATENEGATQVPTSSNRANESQAEQG 60
DB 1 MPASSEKRVHYSTIRKFSVGASVVAASLVMSGVHATENEGATQVPTSSNRANESQAEQG 60
QY 61 EOPKLDSEBDKARKEVEEYVKIIGESYAKSTKRRHTITVALVNLNNIKNEYLINKIVE 120
DB 61 EOPKLDSEBDKARKEVEEYVKIIGESYAKSTKRRHTITVALVNLNNIKNEYLINKIVE 120
QY 121 STSSQLOILMMESRSKYDEAVSKPEKSSSSSSSSSTKPEASDTAKPNKPTBGEKVA 180
DB 121 STSSQLOILMMESRSKYDEAVSKPEKSSSSSSSSSTKPEASDTAKPNKPTBGEKVA 180
QY 181 EAKKVEBAKKAQOKEEDRRNYPTITYKTIETAEBSDEVKKAELVVKANEPDE 240
DB 181 EAKKVEBAKKAQOKEEDRRNYPTITYKTIETAEBSDEVKKAELVVKANEPDE 240
QY 241 OKIQAEEAVESKQAEATRLKKITDREAEBAEKRADAKEOGKPKGRARGVGELAT 300
DB 241 OKIQAEEAVESKQAEATRLKKITDREAEBAEKRADAKEOGKPKGRARGVGELAT 300
QY 301 PDKKENDAKSSDSSVGEETLPSPLKPEKVAEAKKEVKAADQOKEEDRRNYPTNT 360
DB 301 PDKKENDAKSSDSSVGEETLPSPLKPEKVAEAKKEVKAADQOKEEDRRNYPTNT 360
QY 361 YKTIETAEBSDEVKKAELVVKANEPDEKKAELVVKANEPDEKKAELVVKANEPDE 420
DB 361 YKTIETAEBSDEVKKAELVVKANEPDEKKAELVVKANEPDEKKAELVVKANEPDE 420
QY 421 KKAEEBAKKAEEBKVKEKPAEOPAPAPAKPAEKPAPAKPENPAEQKAEKPADQOAE 480
DB 421 KKAEEBAKKAEEBKVKEKPAEOPAPAPAKPAEKPAPAKPENPAEQKAEKPADQOAE 480
QY 481 EDYARSEEEYNRLTQOOPKTEKPAQPSSTP 511
DB 481 EDYARSEEEYNRLTQOOPKTEKPAQPSSTP 511

RESULT 5

US-09-298-523B-12
; Sequence 12, Application US/09298523B
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.

; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; TITLE OF INVENTION: AND STRAINS THEREOF AND USES THEREFOR
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/298,523B
; CURRENT FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-298-523B-12

Query Match 99.6%; Score 2526; DB 16; Length 513;
Best Local Similarity 99.6%; Pred. No. 6,2e-129;
Matches 509; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MPASSEKRVHYSTIRKFSVGASVVAASLVMSGVHATENEGATQVPTSSNRANESQAEQG 60
DB 1 MPASSEKRVHYSTIRKFSVGASVVAASLVMSGVHATENEGATQVPTSSNRANESQAEQG 60
QY 61 EOPKLDSEBDKARKEVEEYVKIIGESYAKSTKRRHTITVALVNLNNIKNEYLINKIVE 120
DB 61 EOPKLDSEBDKARKEVEEYVKIIGESYAKSTKRRHTITVALVNLNNIKNEYLINKIVE 120
QY 121 STSSQLOILMMESRSKYDEAVSKPEKSSSSSSSSSTKPEASDTAKPNKPTBGEKVA 180
DB 121 STSSQLOILMMESRSKYDEAVSKPEKSSSSSSSSSTKPEASDTAKPNKPTBGEKVA 180
QY 181 EAKKVEBAKKAQOKEEDRRNYPTITYKTIETAEBSDEVKKAELVVKANEPDE 240
DB 181 EAKKVEBAKKAQOKEEDRRNYPTITYKTIETAEBSDEVKKAELVVKANEPDE 240
QY 241 OKIQAEEAVESKQAEATRLKKITDREAEBAEKRADAKEOGKPKGRARGVGELAT 300
DB 241 OKIQAEEAVESKQAEATRLKKITDREAEBAEKRADAKEOGKPKGRARGVGELAT 300
QY 301 PDKKENDAKSSDSSVGEETLPSPLKPEKVAEAKKEVKAADQOKEEDRRNYPTNT 360
DB 301 PDKKENDAKSSDSSVGEETLPSPLKPEKVAEAKKEVKAADQOKEEDRRNYPTNT 360
QY 361 YKTIETAEBSDEVKKAELVVKANEPDEKKAELVVKANEPDEKKAELVVKANEPDE 420
DB 361 YKTIETAEBSDEVKKAELVVKANEPDEKKAELVVKANEPDEKKAELVVKANEPDE 420
QY 421 KKAEEBAKKAEEBKVKEKPAEOPAPAPAKPAEKPAPAKPENPAEQKAEKPADQOAE 480
DB 421 KKAEEBAKKAEEBKVKEKPAEOPAPAPAKPAEKPAPAKPENPAEQKAEKPADQOAE 480
QY 481 EDYARSEEEYNRLTQOOPKTEKPAQPSSTP 511
DB 481 EDYARSEEEYNRLTQOOPKTEKPAQPSSTP 511

RESULT 6

US-09-298-523C-12
; Sequence 12, Application US/09298523C
; GENERAL INFORMATION:
; APPLICANT: BRILES et al.
; TITLE OF INVENTION: PNEUMOCOCCAL SURFACE PROTEIN C (PSPC), EPITOPIC REGIONS
; TITLE OF INVENTION: AND STRAINS THEREOF AND USES THEREFOR
; FILE REFERENCE: 454312-3140
; CURRENT APPLICATION NUMBER: US/09/298,523C
; CURRENT FILING DATE: 1999-04-23
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-298-523C-12

Query Match 99.6%; Score 2526; DB 16; Length 513;

Qy	36	YKLTLEIASDVYKKAELTELVEEAKSPREEVKYKAAVSKKAATRLREKITTDR	420
Db	361	YKLTLEIASDVYKKAELTELVEEAKSPREEVKYKAAVSKKAATRLREKITTDR	420
Qy	421	KKAAEEAKRAAEEDYKKEKPAEOPAPAPAEKPAAPKPEMNAEOPKAEKPADQAAE	480
Db	421	KKAAEEAKRAAEEDYKKEKPAEOPAPAPAEKPAAPKPEMNAEOPKAEKPADQAAE	480
Qy	481	EDYARRSEEEYNRLTQOOPKTEKPAQSTP	511
Db	481	EYARRSEEEYNRLTQOOPKTEKPAQSTP	511

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match 99.2%; Score 2516; DB 22; Length 693;

Matches 511; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

Qy	MFPSKSEBRKVHSIRKFSVG-ASVVAVSLIWMGSVVHATENGAIOVPLTSSNRANESQAEO	55
Dp	1 MFPSKSEBRKVHSIRKFSVGASVVAVSLIWMGSVVHATENGAIOVPLTSSNRANESQAEO	60
Qy	60 GEPKPLDSEBRDKARKEVEYKKIVGESYAKSTKKRTTIIVALVNEIANNIKNEYLANKIV	119
Dp	61 GEPKPLDSEBRDKARKEVEYKKIVGESYAKSTKKRTTIIVALVNEIANNIKNEYLANKIV	120
Qy	120 ESTSESOLOLIMMESKSKYDEAVSKFEKODSSSSSDSSTPEASDPAKPKPTPEGEKV	179
Dp	121 ESTSESOLOLIMMESKSKYDEAVSKFEKODSSSSSDSSTPEASDPAKPKPTPEGEKV	180
Qy	180 AEAKKKVVEEA-KKAKQKQKEDRRNPPTTYKYLEIEMESDVEYKKALELVKKANEP	238
Dp	181 AEAKKKVVEEAKKAKQKQKEDRRNPPTTYKYLEIEMESDVEYKKALELVKKANEP	240
Qy	239 DEOKIKOAEAEVESKQAEATRLKKIKTDBEEAEBEAKRADAKEGKPKGARKGVGEL	298
Dp	241 DEOKIKOAEAEVESKQAEATRLKKIKTDBEEAEBEAKRRADAKEGKPKGARKGVGEL	300
Qy	299 ATPDKKENAKSSDSVGBEETLPSPLPKPEKKVMAEKQVVEAKKKADQKQEDRRNPPT	358
Dp	301 ATPDKKENAKSSDSVGBEETLPSPLPKPEKKVMAEKQVVEAKKKADQKQEDRRNPPT	360
Qy	359 NTYKYLEIEMESDVEYKKALELVKEAKEPRNEEKVQKAEVESKKAATRLKIKT	418
Dp	361 NTYKYLEIEMESDVEYKKALELVKEAKEPRNEEKVQKAEVESKKAATRLKIKT	420
Qy	419 DRKKAEEBAKRRKAAEDDKYKEPAQOPAPAPAKAPAPKPRENAEOPAKAPKPADQO	478
Dp	421 DRKKAEEBAKRRKAAEDDKYKEPAQOPAPAPAKAPAPKPRENAEOPAKAPKPADQO	480

```

QY      479 AEDDYARSEEEYNRLTQQPPTKTKPAQPSTP 511
         |||||
Db      481 AEDDYARSEEEYNRLTQQPPTKTKPAQPSTP 513

```

```

RESULT 10
US-10-333-120A-10
Sequence 10, Application US/10333120A
GENERAL INFORMATION:
APPLICANT: Januliczky, Robert
APPLICANT: Iamelli, Francesco
APPLICANT: Sjöholm, Anders
APPLICANT: Pozzi, Gianni
APPLICANT: Björck, Lars
TITLE OF INVENTION: PROTEIN
FILE REFERENCE: 13519PCTUS
CURRENT APPLICATION NUMBER: US/10/333.120A
PRIORITY FILING DATE: 2002-01-14
PRIORITY APPLICATION NUMBER: PCT/EP01/08409
PRIORITY FILING DATE: 2001-07-19
PRIORITY APPLICATION NUMBER: SE 0002728-4
PRIORITY FILING DATE: 2000-07-20
NUMBER OF SEQ ID NOS: 11
SOFTWARE: FaalSEQ for Windows Version 4.0
SEQ ID NO 10
LENGTH: 693
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-10-333-120A-10

```

Query Match	Score	DB	Length
99.2%	2516	29	693

Matches 511; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

1 MFASKSERKVHYSIRKFSVG-ASVVVASLVMGSVHATENEGATQVPTSSNRANESQAEQ 59

Db 1 MFASKSERKVHYSIRKFSVGASVVASLVMGVSVHATENEGATQVPTSSNRANESQAEQ 600

60 GEOPKCLDSERDKARKEVEEYVKIRIGESYAKSTKRRHITVALVNELNINKNEYLNKI V 11

Db 61 GEQPKLDSERDKARKEVEEYVKI VGESYAKSTKKRHTITVALVNELNNIKNEYLNKIV 12

120 ESTSESQLOILMESRSKVDEAVSKFEKSSSSSSSDSTKPEASDTAKPNKPTPEGEKV 17

Db 121 ESTSESQLOILMESRSKVDEAVSKFEKSSSSSSSSDSTKPEASDTAKPNKPTPEGKV 18

QY 180 AEAKKVEEA-KKADQKEEDRNPITYKTLELEIAESDVEVKAELELVKANEP 23

Db 181 AEAKKVEAEKAKDQKEDRRNPTITYKTLELEIAESDVEVKAELELVKANEP 24

QY 239 DEQKIQAEAEVESKQAEATRLKKIKTDREEAEFEAKRRADAKEQGKPKGRAKRGVPGEL 29

Db 241 DEQIKQAEAEVESKQAEATRLKIKTDREEAEBAKRRADAKEQGKPKGRAKRGVPEL 300

QY 299 ATPDKKENDAKSSDSSVGEETLPSPSLKPEKKVAEAEKKEVEEAKKKAAEDQKEEDRRNYPYPT 35

Dd 301 ATPDKKENDAKSSDSSVGEETLPSPSLKPEKKVAEAEKKVEEAKKKAAEDOKEDRRNYP 366

OV 359 NTYKTLLELEIAESDVEVKAELELVKEAKEPNEEKVKOAKAEVESKKAATRIEKI KT 41

Db 361 NTYKTI,ET,IAESDVEYKKA,ET,VKEEAKEPNEEKVYKOKAKAEVESKKARTRI,ET,KT 42

419 DPKKAEFFFAKPKAAEFDKVKKKPAEOPDABADPKAEKPBADABKPBENBAEOPKAEKPADNO 47

431 DPKVEFEVKBKAEEDKIKKEKPAEODORDAPDKVEKBPADKVEENDAEODKAEKPDADCO 48

[illegible]

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

RESULT 11

Sequence 2, Application US/09056019A
GENERAL INFORMATION:
APPLICANT: Tuomanen, Elaine I
APPLICANT: Witzemann, Theresa
APPLICANT: Masure, H. R.
APPLICANT: Johnson, Leslie S.
APPLICANT: Koenig, Scott
TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL
TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED
FILE REFERENCE: 1340-1-017
CURRENT APPLICATION NUMBER: US/09/056, 019A
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 655
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-056-019-2

Query Match 92.6%; Score 2350.5; DB 14; Length 655;
Best Local Similarity 99.8%; Pred. No. 3e-119;
Matches 474; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 38 ENEGATOVPTSSNRANESQAEQEQPKLDSERDKARKEVEYVKIIGESYAKSTKRH 97
DB 1 ENEGATOVPTSSNRANESQAEQEQPKLDSERDKARKEVEYVKIIGESYAKSTKRH 60
QY 98 TITVALVNLNNIKNEYINKIVESTSESQIILMESRSKYDEAVSKFEKSSSSSDS 157
DB 61 TITVALVNLNNIKNEYINKIVESTSESQIILMESRSKYDEAVSKFEKSSSSSDS 120
QY 158 STKPEASDTAKPNKPTREGKVAEAKKVEEA-KKAKQKEEDRNYPITTYKTLELEIA 216
DB 121 STKPEASDTAKPNKPTREGKVAEAKKVEEA-KKAKQKEEDRNYPITTYKTLELEIA 180
QY 217 ESDVEVKKAELVYKVAANEPRDEQIKQAEVESKQAEATRLKKTIDREAEAEBAKR 276
DB 181 ESDVEVKKAELVYKVAANEPRDEQIKQAEVESKQAEATRLKKTIDREAEAEBAKR 240
QY 277 RADAKEQCKPKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPSPSLKPEKKVAEAK 336
DB 241 RADAKEQCKPKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPSPSLKPEKKVAEAK 300
QY 337 KVEEAKKKAEDQEKEDRNYPITTYKTLELEIAESDVYKKAELVYKAEKEPRNEBK 396
DB 301 KVEEAKKKAEDQEKEDRNYPITTYKTLELEIAESDVYKKAELVYKAEKEPRNEBK 360
QY 397 KOAKAEVSKKAETRLKKTIDRKKAEEBAKRAEEDVKYKEKPAOPAPAPAKAEKP 456
DB 361 KOAKAEVSKKAETRLKKTIDRKKAEEBAKRAEEDVKYKEKPAOPAPAPAKAEKP 420
QY 457 APAPKPNPAPQPKAEKPADQAEEDYARSESEYNRLTQOQPPKTEKPAQPSPT 511
DB 421 APAPKPNPAPQPKAEKPADQAEEDYARSESEYNRLTQOQPPKTEKPAQPSPT 475

RESULT 12
US-09-056-019B-2
Sequence 2, Application US/09056019B
GENERAL INFORMATION:
APPLICANT: Tuomanen, Elaine I
APPLICANT: Witzemann, Theresa M.
APPLICANT: Masure, H. R.
APPLICANT: Johnson, Leslie S.
APPLICANT: Koenig, Scott
TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL
TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED
FILE REFERENCE: 5853-2
CURRENT APPLICATION NUMBER: US/09/056, 019B
CURRENT FILING DATE: 1998-04-07

NUMBER OF SEQ ID NOS: 40
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 655
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-056-019B-2

Query Match 92.6%; Score 2350.5; DB 14; Length 655;
Best Local Similarity 99.8%; Pred. No. 3e-119;
Matches 474; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 38 ENEGATOVPTSSNRANESQAEQEQPKLDSERDKARKEVEYVKIIGESYAKSTKRH 97
DB 1 ENEGATOVPTSSNRANESQAEQEQPKLDSERDKARKEVEYVKIIGESYAKSTKRH 60
QY 98 TITVALVNLNNIKNEYINKIVESTSESQIILMESRSKYDEAVSKFEKSSSSSDS 157
DB 61 TITVALVNLNNIKNEYINKIVESTSESQIILMESRSKYDEAVSKFEKSSSSSDS 120
QY 158 STKPEASDTAKPNKPTREGKVAEAKKVEEA-KKAKQKEEDRNYPITTYKTLELEIA 216
DB 121 STKPEASDTAKPNKPTREGKVAEAKKVEEA-KKAKQKEEDRNYPITTYKTLELEIA 180
QY 217 ESDVEVKKAELVYKVAANEPRDEQIKQAEVESKQAEATRLKKTIDREAEAEBAKR 276
DB 181 ESDVEVKKAELVYKVAANEPRDEQIKQAEVESKQAEATRLKKTIDREAEAEBAKR 240
QY 277 RADAKEQCKPKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPSPSLKPEKKVAEAK 336
DB 241 RADAKEQCKPKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPSPSLKPEKKVAEAK 300
QY 337 KVEEAKKKAEDQEKEDRNYPITTYKTLELEIAESDVYKKAELVYKAEKEPRNEBK 396
DB 301 KVEEAKKKAEDQEKEDRNYPITTYKTLELEIAESDVYKKAELVYKAEKEPRNEBK 360
QY 397 KOAKAEVSKKAETRLKKTIDRKKAEEBAKRAEEDVKYKEKPAOPAPAPAKAEKP 456
DB 361 KOAKAEVSKKAETRLKKTIDRKKAEEBAKRAEEDVKYKEKPAOPAPAPAKAEKP 420
QY 457 APAPKPNPAPQPKAEKPADQAEEDYARSESEYNRLTQOQPPKTEKPAQPSPT 511
DB 421 APAPKPNPAPQPKAEKPADQAEEDYARSESEYNRLTQOQPPKTEKPAQPSPT 475

RESULT 13
US-09-287-788-2
Sequence 2, Application US/09287788B
GENERAL INFORMATION:
APPLICANT: Masure, H. R.
APPLICANT: Johnson, Leslie S.
TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL
TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED
FILE REFERENCE: 1340-1-017N
CURRENT APPLICATION NUMBER: US/09/287, 788B
CURRENT FILING DATE: 1999-04-07
PRIOR APPLICATION NUMBER: 60/080, 878
PRIOR FILING DATE: 1998-04-07
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 655
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-287-788-2

Query Match 92.6%; Score 2350.5; DB 16; Length 655;
Best Local Similarity 99.8%; Pred. No. 3e-119;
Matches 474; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 38 ENEGATOVPTSSNRANESQAEQEQPKLDSERDKARKEVEYVKIIGESYAKSTKRH 97

Db 1 ENEGATVPTSSNRANESQAEQEQPKLDSERDARKEVEYKKIVGESYAKSTKKRH 60
Qy 98 TTTVALVNELNNIKNEYLINKIVESTSESOLIMMESRSKVDEAVSKFEKSSSSSSDS 157
Db 61 TTTVALVNELNNIKNEYLINKIVESTSESOLIMMESRSKVDEAVSKFEKSSSSSSDS 120
Qy 158 STKPEASDTAKPNKPTPEGEKVAEAKKVEEA-KKADQKEEDRRNPTTYKTLELEIA 216
Db 121 STKPEASDTAKPNKPTPEGEKVAEAKKVEEA-KKADQKEEDRRNPTTYKTLELEIA 180
Qy 217 ESDVEVKAELELVKVAKNPDEOKIKQAEAVESKQAEATRLKKITDSEEAEEKR 276
Db 181 ESDVEVKAELELVKVAKNPDEOKIKQAEAVESKQAEATRLKKITDSEEAEEKR 240
Qy 217 RADAEQKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPSLKPEKKVAEAK 336
Db 241 RADAEQKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPSLKPEKKVAEAK 300
Qy 337 KVEAKKKAEDQKEEDRRNPTTYKTLELEIAESDVVKAELELVKEAEKPEPNEEKV 396
Db 301 KVEAKKKAEDQKEEDRRNPTTYKTLELEIAESDVVKAELELVKEAEKPEPNEEKV 360
Qy 397 KQAKAEVSKQAEATRLKIKTRKKAEEBAKRAAEEDKYKEKPAEQPPAPAKAEK 456
Db 361 KQAKAEVSKQAEATRLKIKTRKKAEEBAKRAAEEDKYKEKPAEQPPAPAKAEK 420
Qy 457 APAPKPPAPKPAKPAEQPPAPAKAEEDVARSSEENRLLTQOOPPTTEKPAQSP 511
Db 421 APAPKPPAPKPAKPAEQPPAPAKAEEDVARSSEENRLLTQOOPPTTEKPAQSP 475

RESULT 14

US-09-056-019-38

Sequence 38, Application US/09056019A
GENERAL INFORMATION:
APPLICANT: Tuomanen, Elaine I
APPLICANT: Wizemann, Theresa
APPLICANT: Maure, H. R.
APPLICANT: Johnson, Leslie S.
APPLICANT: Koenig, Scott
TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL
TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED
FILE REFERENCE: 1340-1-017
CURRENT APPLICATION NUMBER: US/09/056, 019A
CURRENT FILING DATE: 1998-04-07
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 38
LENGTH: 460
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-056-019-38

Query Match 88.8%; Score 2252.5; DB 14; Length 460;

Best Local Similarity 99.6%; Pred. No. 4.1e-114; Indels 1; Gaps 1;

Matches 457; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

Qy 24 VVASLVGSVVHATENEGATVPTSSNRANESQAEQEQPKLDSERDARKEVEYKK 83
Db 2 IVASLVGSVVHATENEGATVPTSSNRANESQAEQEQPKLDSERDARKEVEYKK 61
Qy 84 IVGESYAKSTKKRTTTTVALVNELNNIKNEYLINKIVESTSESOLIMMESRSKVDEAVS 143
Db 62 IVGESYAKSTKKRTTTTVALVNELNNIKNEYLINKIVESTSESOLIMMESRSKVDEAVS 121
Qy 144 KFEKSSSSSSSDSTKPEASDTAKPNKPTPEGEKVAEAKKVEEA-KKADQKEEDRRN 202
Db 122 KFEKSSSSSSSDSTKPEASDTAKPNKPTPEGEKVAEAKKVEEA-KKADQKEEDRRN 181
Qy 203 YPTTYKTLELEIAESDVVKAELELVKVAKNPDEOKIKQAEAVESKQAEATRLKK 262
Db 182 YPTTYKTLELEIAESDVVKAELELVKVAKNPDEOKIKQAEAVESKQAEATRLKK 241

Qy 263 IKTDREAEBAKRRADAKQEGKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPS 322
Db 242 IKTDREAEBAKRRADAKQEGKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPS 301
Qy 323 PSLKPEKKVAEAKKVEEA-KKADQKEEDRRNPTTYKTLELEIAESDVVKAELEL 382
Db 302 PSLKPEKKVAEAKKVEEA-KKADQKEEDRRNPTTYKTLELEIAESDVVKAELEL 361
Qy 383 VKERAEKPEPNEEKIKQAEAVESKQAEATRLKKITDREAEBAKRRADAKQEGKPK 442
Db 362 VKERAEKPEPNEEKIKQAEAVESKQAEATRLKKITDREAEBAKRRADAKQEGKPK 421
Qy 443 EQPPAPAKPAKPAEQPPAPAKAEEDVARSSEENRLLTQOOPPTTEKPAQSP 511
Db 422 EQPPAPAKPAKPAEQPPAPAKAEEDVARSSEENRLLTQOOPPTTEKPAQSP 475

RESULT 15

US-09-056-019B-38

Sequence 38, Application US/09056019B
GENERAL INFORMATION:
APPLICANT: Tuomanen, Elaine I
APPLICANT: Wizemann, Theresa M.
APPLICANT: Maure, H. R.
APPLICANT: Johnson, Leslie S.
APPLICANT: Koenig, Scott
TITLE OF INVENTION: POLYPEPTIDE COMPRISING THE AMINO ACID OF AN N-TERMINAL
TITLE OF INVENTION: CHOLINE BINDING PROTEIN A TRUNCATE, VACCINE DERIVED
FILE REFERENCE: 5853-2
CURRENT APPLICATION NUMBER: US/09/056, 019B
CURRENT FILING DATE: 1998-04-07
NUMBER OF SEQ ID NOS: 40
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 38
LENGTH: 460
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-056-019B-38

Query Match 88.8%; Score 2252.5; DB 14; Length 460;

Best Local Similarity 99.6%; Pred. No. 4.1e-114; Indels 1; Gaps 1;

Matches 457; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

Qy 24 VVASLVGSVVHATENEGATVPTSSNRANESQAEQEQPKLDSERDARKEVEYKK 83
Db 2 IVASLVGSVVHATENEGATVPTSSNRANESQAEQEQPKLDSERDARKEVEYKK 61
Qy 84 IVGESYAKSTKKRTTTTVALVNELNNIKNEYLINKIVESTSESOLIMMESRSKVDEAVS 143
Db 62 IVGESYAKSTKKRTTTTVALVNELNNIKNEYLINKIVESTSESOLIMMESRSKVDEAVS 121
Qy 144 KFEKSSSSSSSDSTKPEASDTAKPNKPTPEGEKVAEAKKVEEA-KKADQKEEDRRN 202
Db 122 KFEKSSSSSSSDSTKPEASDTAKPNKPTPEGEKVAEAKKVEEA-KKADQKEEDRRN 181
Qy 203 YPTTYKTLELEIAESDVVKAELELVKVAKNPDEOKIKQAEAVESKQAEATRLKK 262
Db 182 YPTTYKTLELEIAESDVVKAELELVKVAKNPDEOKIKQAEAVESKQAEATRLKK 241
Qy 263 IKTDREAEBAKRRADAKQEGKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPS 322
Db 242 IKTDREAEBAKRRADAKQEGKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPS 301
Qy 323 PSLKPEKKVAEAKKVEEA-KKADQKEEDRRNPTTYKTLELEIAESDVVKAELEL 382
Db 302 PSLKPEKKVAEAKKVEEA-KKADQKEEDRRNPTTYKTLELEIAESDVVKAELEL 361
Qy 383 VKERAEKPEPNEEKIKQAEAVESKQAEATRLKKITDREAEBAKRRADAKQEGKPK 442
Db 362 VKERAEKPEPNEEKIKQAEAVESKQAEATRLKKITDREAEBAKRRADAKQEGKPK 421

Mon Nov 24 10:00:24 2003

us-09-298-523c-13.رام

Page 8

Dy 443 EQPQAPAPKAKEPPAPAPKPPENPAEQPKAEKPADQAEE 481
Dd 422 EQPQAPAPKAKEPPAPAPKPPENPAEQPKAEKPADQAEE 460

Search completed: November 21, 2003, 13:40:57
Job time : 247.996 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw.model

Run on: November 21, 2003, 13:29:32 ; Search time 20.7062 Seconds
(without alignment)
2373.306 Million cell updates/sec

Title: US-09-298-523C-13

Perfect score: 2537

Sequence: 1 MFASKSRKRVHYGIRKPSVG.....NRLTQQQPPKTEKPAQSTP 511

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: PIR.76:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2516	99.2	693	2	H95255 choline binding pr
2	1730	68.2	701	2	H98120 choline binding pr
3	491	19.4	619	2	A41971 surface protein ps
4	491	19.4	619	2	A97887 surface protein ps
5	397.5	15.7	744	2	P95013 pneumococcal surfa
6	389.5	15.4	1110	2	I51116 NF-180 - sea lampr
7	363	14.3	1020	1	QPHUH neurofilament trip
8	361	14.2	1701	2	T09127 probable erythrocy
9	351.5	13.9	1164	1	FC50AG IGA Fc receptor pr
10	346.5	13.7	1134	2	A60234 IGA Fc receptor pr
11	345	13.6	606	2	A43477 neurofilament trip
12	337.5	13.3	1052	2	A44937 kineoclast-associ
13	329.5	13.0	3488	2	T34418 hypochetrical prote
14	328	12.9	1130	2	S51364 sperm tail-specifi
15	328	12.9	6642	2	T29757 protein UNC-89 - C
16	322	12.7	1072	1	A37221 neurofilament trip
17	318	12.5	854	2	S02003 neurofilament trip
18	315.5	12.4	706	2	A45980 junctional sarcopl
19	315	12.4	522	2	C96608 hypochetrical prote
20	313.5	12.4	699	2	E84565 hypochetrical prote
21	311	12.3	539	2	A28549 M24 protein precur
22	310	12.2	1087	1	QFM5H neurofilament trip
23	308	12.1	5327	2	T13564 microtubule-associ
24	300	11.8	771	1	A33430 h-caldesmon - chic
25	293	11.5	1526	2	A43605 mature-parasite-in
26	292.5	11.5	858	2	S15762 neurofilament trip
27	292.5	11.5	1815	2	C81169 IGA-specific metal
28	292	11.5	1094	2	S49313 protein kinase - s
29	290	11.4	729	2	S68191 triadin - human

30	288	11.4	1128	2	T30296 R27-2 protein - Tr
31	285.5	11.3	501	2	A44643 M protein precursor
32	285.5	11.3	1773	2	A81937 IGA-specific metal
33	284	11.2	644	2	S55395 neurofilament prot
34	283	11.2	916	2	A27864 neurofilament trip
35	282.5	11.1	849	2	S00030 neurofilament trip
36	279	11.0	667	2	A40713 cylicin II - bovin
37	278	11.0	488	2	I46014 cylicin II - bovin
38	275	10.8	1192	2	A71623 probable secreted
39	274.5	10.8	332	2	B43427 neurofilament prot
40	274	10.8	2464	1	QRMSP1 microtubule-associ
41	270.5	10.7	598	2	B40713 cylicin I - human
42	269.5	10.6	532	2	S54871 M protein - Strept
43	268.5	10.6	798	2	I50479 neurofilament medi
44	268.5	10.6	2364	2	A56577 microtubule-associ
45	266	10.5	483	2	A26297 M6 protein - Strept

ALIGNMENTS

RESULT 1					
H95255 choline binding protein A [imported] - Streptococcus pneumoniae (strain TIGR4)					
C/Species: Streptococcus pneumoniae					
C/Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001					
C/Accession: H95255					
R/tetelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid, on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzaple, nson, T.; Hickey, E.K.; Holt, I.E.					
A/Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison, A.; Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.					
A/Reference number: A95000; MUID:21357209; PMID:11463916					
A/Accession: H95255					
A/Status: preliminary					
A/Molecule type: DNA					
A/Residues: 1-693 <KIR>					
A/Cross-references: GB:AE005672; PIDN:AAK76241.1; PID:G14973701; GSPDB:GN00164; TIGR:SP4					
A/Experimental source: strain TIGR4					
C/Genetics:					
A/Gene: SP2190					
Query Match					
Best Local Similarity 99.2%; Score 2516; DB 2; Length 693;					
Matches 511; Conservative 0; Mismatches 0; Indels 2; Gaps 2;					
QY	1	MFASKSRKRVHYGIRKPSVG-ASVVVASLVGSVVHATENEGATQVPTSSNRANESQAQ	59		
DB	1	MFASKSRKRVHYGIRKPSVGASVVVASLVGSVVHATENEGATQVPTSSNRANESQAQ	60		
QY	60	GEOPKLDSSRDARKEVEEYVKKIVGESYAKSTKRRHTTVALVNLNNIKNYEYLNKIV	119		
DB	61	GEOPKLDSSRDARKEVEEYVKKIVGESYAKSTKRRHTTVALVNLNNIKNYEYLNKIV	120		
QY	120	ESTSSESOQLIMMESSKVDVAISKPEKDSSSSSSSSTKPEASDPKANKPKPEBEKV	179		
DB	121	ESTSSESOQLIMMESSKVDVAISKPEKDSSSSSSSSTKPEASDPKANKPKPEBEKV	180		
QY	180	AEKKKVEEA-KKAKOKEEDRRNYPITYKTLELEIAESDVEYKAELELVKANEPR	238		
DB	181	AEKKKVEEA-KKAKOKEEDRRNYPITYKTLELEIAESDVEYKAELELVKANEPR	240		
QY	239	DEQIKQAEAEVSKQAEATRLKIKITDRBEAEBAARRADAEQPKGRAKRGVQEL	298		
DB	241	DEQIKQAEAEVSKQAEATRLKIKITDRBEAEBAARRADAEQPKGRAKRGVQEL	300		
QY	299	ATDDKKENDAKSSDSSVGERTLPSPSLKPEKTAVAEAKKYAEAKKKAEDQKEDRRNYP	358		
DB	301	ATDDKKENDAKSSDSSVGERTLPSPSLKPEKTAVAEAKKYAEAKKKAEDQKEDRRNYP	360		
QY	359	NTKTELELEIAESDVEYKAELELVKAEKPEENEEVKQAKAEVSKKAEATRLKIKT	418		

Qy	Db	Qy	Db
419	361	479	481
DRKKKEEAKRAAEEDKYKEKPAEOPAPAKAEKPA	NTYKLELELIAESDVEVKKAELELYEEAKEPENEEKVQAKKEVSS	AEEDYARSRSEENRLTOOQPPKTEKPAOPSTP	AEEDYARSRSEENRLTOOQPPKTEKPAOPSTP
421	419	511	513
DRKKEEAKRAAEEDKYKEKPAEOPAPAKAEKPA	NTYKLELELIAESDVEVKKAELELYEEAKEPENEEKVQAKKEVSS	AEEDYARSRSEENRLTOOQPPKTEKPAOPSTP	AEEDYARSRSEENRLTOOQPPKTEKPAOPSTP

RESULT 2

cholesterol binding protein A [imported] - Streptococcus pneumoniae (strain R6)
C/Species: Streptococcus pneumoniae
C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C/Accession: H98120
R/Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczyk, L.; Burgelt, S.; DeHoff, B.S.; F
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; N
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A/Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.
A/Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A/Reference number: A97872; MUID:2143245; PMID:11544234
A/Accession: H98120
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-701 <KUP>
A/Cross-references: GB:AE007317, PID:AA00797.1, PID:G15459699, GSEDB:GN00174
A/Genetics:
A/Gene: pspC

Query Match	68.2%	Score 1730	DB 2	length 701
Best Local Similarity	70.7%	Pred. No. 1.6e-59		
Matches 371	Conservative 49	Mismatches 47	Indels 58	Gaps 10

```

OY      1 MPAKSERKHYSTIRKFSV-ASVVASLWMSVHATENGCAOPTSNNRANESQAQ 59
DB      1 MPAKSERKHYSTIRKFSIGVASVAVASLWMSGVHATENEGSTQATSSMAK----- 54

```

```
0Y      60 GEOPKLDSEBDKARKE-VEEYVKIKVIGESYAKSTKKRRHITVALVNELNNINXEYLNK-    117
        : | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
0b -----TEBUBAAKAOIVDEYIETEMT BF -TOL DPKPCTONVANLNIKRI SA IETVKYI BBT     104
55 -----
```

```
QY      118 -I V E S T S E S Q I L M M E R S K V D E A V S K F E K U D S S S S S S D S T K P E A S D T A K E N K P T E R G   176
```

QY	177	EKAFAKKUYEEA-KKAKOKEEDRRNPTTYKTLEIEIAESDVEVVKAELEIVRVKAN	235
22	103	NATLADICACUDEF--GELINADDPRTENFINDI-----LNFG	138

Dd 139 EKVAEAKKKVEAKKKKADQKEEDRRNYPNTYKTLLELEIAEFDVYKAELELVEAK 198

```
Db      199 ESNEGTIKAKEVESKKAETRIJENIKTRKKAEEBZAKADAKLKEANVAATSDQGP 258
```

Db
259 KGRAKRGVPELATPPDKENDAKSSDSSYGEEETLPSSSLKSGKKVAEAEKKYVEAEKKRX 318

Qy 347 DQKEDRRNPNTYKTELEIAESDVEVKKALELVEEAKEPNEEKVKQAKAEVESK 406

OY 407 KAAATRLKIKTDKKAABEAKRKAABEDKYKEKPAEQOPAPAPAKAEKPPAKPPENDA 466

Db 379 KAEATRLINIKTDKKAAEEBAKRRKAAEEDKYKKEPAEQQPAPAPQPEK--PAKPKEKA 436

D5 437 EDPKAEKTDQAEEDYARRSEEEYNRLTQQPDKTEKPAQPSIP 481

RESULT 3
A41971

C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 15-Oct-1999
C:Accession: A41971, A60287, A33134

A:Title: Structural properties and evolutionary relationships of PcpA, a surface protein
A:Reference number: A41971; MUID:92105030; PMID:1729249
A:Accession number: A41971

```

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-619 <YOT>

```

ANote: sequence extracted from NCBI backbone (NCBIN:75635, NCBI:75636)
R.Talkingdon, D.F.; Citimmins, D.L.; Voellinger, D.C.; Yether, J.; Briles, D.E.
Infect. Immun. 59, 1285-1289, 1991

A:Reference number: A60282; MWID:91169598; PMID:2004810
A:Accession: A60282
A:Molecule type: protein

A;Experimental source: strain JY2008
C;Genetics:
A;Gene: pspA

F;1-31/Domain: signal sequence #status predicted <SIG>
F;32-619/Product: surface protein pspa #status predicted <MAT>
F;411-430/Domain: cpl repeat homology <CP01>

F;451-470/Domain:	cpl repeat	homology	<CP03>
F;471-490/Domain:	cpl repeat	homology	<CP04>
F;491-510/Domain:	cpl repeat	homology	<CP05>

F;531-550/Domain:	cpl repeat	homology	<CP07>
F;551-570/Domain:	cpl repeat	homology	<CP08>
F;571-591/Domain:	cpl repeat	homology	<CP09>

Query Match	Score	DB 2:	Length
19.4%	491	619	
37.3%	70	13	

161	PEASDTAKBNKPTBGEKVAE-AKKKVEEAKKAKKADKKKEDRRNVPITYKTI,FI,ETAE	SD 219
Matches	152; Conservative	50; Mismatches 109; Indels 96; Gaps 18;

Db 35 PVASQ : SIAEKDYDAKKDAKNAKKAKEDEQAALBDADAKAAQAQKYDEDEDQKTEE----- 86

```

      87  ---KALEKA---ASEEMDKVAVAQQAYLAYQCATD---KAAKDADADSIDAEKKR-- 134

```

QY 280 AKEQGKPRGAKRGV----PGLATPKDKENDAKSSDSVGEBTLPSPLKPEKVAEAE 335
:::|::|::|:
Db :|:|:|:
135 -DEAYKMEENTVPAMTIDDPDPTLSTKYVCSEALVGV-----APR-TPYLFPKK 391

QY 336 KVEEAKKAAD--QKEDRNYP-----TNTYKTLLELEAESD----- 372

00 -----VEYKKAELIVKE-----FAKEPRNEEKVOKAAE-----VES 405
 102 ANUBERENNAI BRANAVURNEVE VKNIMEDUBINY VKNUBVEJUNELDESBBUDIMNGF RA 241
 373 -----VEYKKAELIVKE-----FAKEPRNEEKVOKAAE-----VES 405

Db 242 PLQSKLPAKKAKLSKLEELSPKIDELDAEIAKLEBQKKAEEENNVEDYFKEGLEKTIAA 301

```

Db
302 KKAEE---LEKTEADLKKAVNBEKPARAPETPARAPAEQKRPAPAP---QPARAPREK 355

```

465 PAEQPKAEKPADQQAEDDYARRSEEEYNRLLTQQQPKTEKPA-QPST 510

22

RESULT 4

surface protein pspA precursor [improved] - Streptococcus pneumoniae (strain R6)
C/Species: Streptococcus pneumoniae
C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C/Accession: A97887
R/Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.; F
R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAnren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A./Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A./Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A./Reference number: A978872; MUID:21429245; PMID:11544224
A/Accession: A97887
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-619 <KRR>
A/Cross-references: GB:AE007317; PIDN:PAK98925.1; PID:915457660; GSPDB:GN00174
C/Genetics:
C./Gene: pspA

A;Experimental source: strain TIGR4
C;Genetics:

Query Match	15.7%;	Score 397.5;	DB 2;	Length 744;
Best Local Similarity	28.0%;	Pred. No. 1.2e-08;		
Matches 147;	Conservative 89;	Mismatches 166;	Indels 123;	Gaps 22

A;Experimental source: strain TIGR4
C;Genetics:

Query Match	15.7%;	Score 397.5;	DB 2;	Length 744;
Best Local Similarity	28.0%;	Pred. No. 1.2e-08;		
Matches 147;	Conservative 89;	Mismatches 166;	Indels 123;	Gaps 22

```

Oy 108 NNKKVEYJNKTVESSTSESOL0---LMMESSKVDVAVSKEKODSSSSSDSSSKPPAS 164
Db 544 EETEAEAAEBEEBAEGEEBAEAGEEAAEEBAVEEBALEKAAEAAKAEVEEBAEABEE 603
Oy 165 DTAKPNKTEPGEKVAAEAKKV-----EAKKAKODKEEDRRNYPIT--YKTTLE 214
Db 604 E-----EEBAEEEEVAETKEEVEABAEVBEEGEAAEEBAEEBAEEBVTSKAKTQAE 659
Oy 215 IASDVVEKKKELEJ-----YKKANEPNDEQIKQAEAVESEKQAEARLKKI- 263
Db 660 VEEBAEAEABAEBAEBAEGEEDVEAESK-EEEBOSKADAEADAEBAEEBVEBVT 718
Oy 264 KTDREAEAEBAEKRRADAKEOGKPKGRARKGVPGELATDPKKENAKSSDSSGEBETLSP 323
Db 719 KSDAEBAEABAEBAE-AKSEBAEABEAKODABEEBAEBAEVAETBAATEBAEAKCA--SD 775
Oy 324 SLKPEKKVAEAEKKVVEAKKKAEDOKEDRRNYPNTYKTTLEIASDVVEKKAELELV 383
Db 776 DEKEEBEYKESBAPAPAEKAKAPEKKAAPKKAAPAK---VESPTSEDEBPK--ABVV 828
Oy 384 KEAEKPEPNBEKVQAKAEVESKKAEATRLTEKIKTRKKAESEAKRKAABEDKVAEKPAE 443
Db 829 EKKGALEAPKPKAKPAALAKKEAKPYE---KEEBEESPTBEPKPPA-----AKPAK 878
Oy 444 QP-QPAPAPKA---EKPAPA---PKPENPAEQPAEKPRADQQAEDY-----483
Db 879 APAPKPAKPAKEABEKKPEPAKPAQKAPAAABEEDKEEDBEEBEVEEYKVEDAKPVK 938
Oy 484 ---ARSEEVNRLTQOOPPKTE--KPAP 508
Db 939 SKPAKAEEDBPKPAKQPPKPKRRPAP 967

```

RESULT 7
 OPHUH
 neurofilament triplet H protein - human
 N:Alternate names: neurofilament protein, 112K
 C:Species: Homo sapiens (man)
 C:Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 16-Jun-2000
 C:Accession: S00979
 R:Lees, J.F.; Shreidman, P.S.; Skuntz, S.F.; Carden, M.J.; Lazzarini, R.A.
 EMBL J. 7, 1947-1955, 1988
 A:Title: The structure and organization of the human heavy neurofilament subunit (NF-H)
 A:Reference number: S00979; MUID:88328981; PMID:3138108
 A:Accession: S00979
 A:Molecule type: DNA
 A:Residues: 1-1020 <LEB>
 A:Cross-references: EMBL:X15306; NID:q35028; PIDN:CAA3366.1; PID:g1841430
 A:Note: It is uncertain whether Met-1 or Met-2 is the initiator
 C:Genetics:
 A:Gene: GDB:NEFH
 A:Cross-references: GDB:120225; OMIM:162230
 A:Map position: 22q12.1-22q13.1
 A:Introns: 295/1; 361/3; 403/2
 C:Superfamily: neurofilament triplet H protein
 C:Keywords: coiled coil; heterotrimer; intermediate filament; nerve; phosphoprotein
 F:1-100/Domain: amino-terminal <NTD>
 F:101-410/Domain: rod #status predicted <ROD>
 F:411-1020/Domain: carboxyl-terminal <CTD>
 F:502-826/Region: 14-residue repeats
 F:503, 511, 518, 526, 532, 540, 546, 552, 560, 566, 574, 580, 586, 594, 600, 606, 614, 620, 628, 634, 640, 646, 650, 656, 660, 666, 670, 676, 682, 688, 694, 698, 704, 710, 716, 722, 728, 734, 740, 746, 752, 758, 764, 770, 776, 782, 788, 794, 800, 806, 812, 818, 824, 830, 836, 842, 848, 854, 860, 866, 872, 878, 884, 890, 896, 902, 908, 914, 920, 926, 932, 938, 944, 950, 956, 962, 968, 974, 980, 986, 992, 998, 1004, 1010, 1016, 1020, 1026, 1032, 1038, 1044, 1050, 1056, 1062, 1068, 1074, 1080, 1086, 1092, 1098, 1104, 1110, 1116, 1122, 1128, 1134, 1140, 1146, 1152, 1158, 1164, 1170, 1176, 1182, 1188, 1194, 1200, 1206, 1212, 1218, 1224, 1230, 1236, 1242, 1248, 1254, 1260, 1266, 1272, 1278, 1284, 1290, 1296, 1302, 1308, 1314, 1320, 1326, 1332, 1338, 1344, 1350, 1356, 1362, 1368, 1374, 1380, 1386, 1392, 1398, 1404, 1410, 1416, 1422, 1428, 1434, 1440, 1446, 1452, 1458, 1464, 1470, 1476, 1482, 1488, 1494, 1500, 1506, 1512, 1518, 1524, 1530, 1536, 1542, 1548, 1554, 1560, 1566, 1572, 1578, 1584, 1590, 1596, 1602, 1608, 1614, 1620, 1626, 1632, 1638, 1644, 1650, 1656, 1662, 1668, 1674, 1680, 1686, 1692, 1698, 1704, 1710, 1716, 1722, 1728, 1734, 1740, 1746, 1752, 1758, 1764, 1770, 1776, 1782, 1788, 1794, 1800, 1806, 1812, 1818, 1824, 1830, 1836, 1842, 1848, 1854, 1860, 1866, 1872, 1878, 1884, 1890, 1896, 1902, 1908, 1914, 1920, 1926, 1932, 1938, 1944, 1950, 1956, 1962, 1968, 1974, 1980, 1986, 1992, 1998, 2004, 2010, 2016, 2022, 2028, 2034, 2040, 2046, 2052, 2058, 2064, 2070, 2076, 2082, 2088, 2094, 2100, 2106, 2112, 2118, 2124, 2130, 2136, 2142, 2148, 2154, 2160, 2166, 2172, 2178, 2184, 2190, 2196, 2202, 2208, 2214, 2220, 2226, 2232, 2238, 2244, 2250, 2256, 2262, 2268, 2274, 2280, 2286, 2292, 2298, 2304, 2310, 2316, 2322, 2328, 2334, 2340, 2346, 2352, 2358, 2364, 2370, 2376, 2382, 2388, 2394, 2400, 2406, 2412, 2418, 2424, 2430, 2436, 2442, 2448, 2454, 2460, 2466, 2472, 2478, 2484, 2490, 2496, 2502, 2508, 2514, 2520, 2526, 2532, 2538, 2544, 2550, 2556, 2562, 2568, 2574, 2580, 2586, 2592, 2598, 2604, 2610, 2616, 2622, 2628, 2634, 2640, 2646, 2652, 2658, 2664, 2670, 2676, 2682, 2688, 2694, 2700, 2706, 2712, 2718, 2724, 2730, 2736, 2742, 2748, 2754, 2760, 2766, 2772, 2778, 2784, 2790, 2796, 2802, 2808, 2814, 2820, 2826, 2832, 2838, 2844, 2850, 2856, 2862, 2868, 2874, 2880, 2886, 2892, 2898, 2904, 2910, 2916, 2922, 2928, 2934, 2940, 2946, 2952, 2958, 2964, 2970, 2976, 2982, 2988, 2994, 3000, 3006, 3012, 3018, 3024, 3030, 3036, 3042, 3048, 3054, 3060, 3066, 3072, 3078, 3084, 3090, 3096, 3102, 3108, 3114, 3120, 3126, 3132, 3138, 3144, 3150, 3156, 3162, 3168, 3174, 3180, 3186, 3192, 3198, 3204, 3210, 3216, 3222, 3228, 3234, 3240, 3246, 3252, 3258, 3264, 3270, 3276, 3282, 3288, 3294, 3300, 3306, 3312, 3318, 3324, 3330, 3336, 3342, 3348, 3354, 3360, 3366, 3372, 3378, 3384, 3390, 3396, 3402, 3408, 3414, 3420, 3426, 3432, 3438, 3444, 3450, 3456, 3462, 3468, 3474, 3480, 3486, 3492, 3498, 3504, 3510, 3516, 3522, 3528, 3534, 3540, 3546, 3552, 3558, 3564, 3570, 3576, 3582, 3588, 3594, 3600, 3606, 3612, 3618, 3624, 3630, 3636, 3642, 3648, 3654, 3660, 3666, 3672, 3678, 3684, 3690, 3696, 3702, 3708, 3714, 3720, 3726, 3732, 3738, 3744, 3750, 3756, 3762, 3768, 3774, 3780, 3786, 3792, 3798, 3804, 3810, 3816, 3822, 3828, 3834, 3840, 3846, 3852, 3858, 3864, 3870, 3876, 3882, 3888, 3894, 3900, 3906, 3912, 3918, 3924, 3930, 3936, 3942, 3948, 3954, 3960, 3966, 3972, 3978, 3984, 3990, 3996, 4002, 4008, 4014, 4020, 4026, 4032, 4038, 4044, 4050, 4056, 4062, 40

[illegible]

OY	60	GEQPKJUDSEDKD--RVEVEYVK-KIVGSGYAKSTKTRHTITITVALVNEIINIINEXYIN	116
Dd	497	GEETKSPPEAEAAKSPEKEAKSPVKEEAKSPAIAKSPKEKEAKSPAIVKSPBEKAKSP--	553
OY	117	KIVESTESQJQIIMMESRSKVYDAVSKFEEDOSSSSSDSSTYPERASDTAKPNKPT--EP	175
Dd	554	--AAKEEAKSPPEAKSPKEEAKSPAIVKSPBEKAKSPAIAKSPAIAKSPAIAKSPAIVKEE	611
OY	176	GEKVAEAAKKKY--EEAKTKAKDQKEEDRRNYPTITYKTTLEIAESDVEVKKAELELVKYKA	234
Dd	612	AKSPAIAEAKSPVKEEAKSPAIVKSPBEKAKSP-----KEAKSPBEKAKSPB-----KEEA	660
OY	235	NEPDEOKIKOAEAEVSK-----QAEATRLKIKITD--REBAEBAKRRADAKEOGKPKG	288
Dd	661	KSPBEKAKSPAIVAEAKSPBEKAKSPAIVAEAKSPBEKAKSPAIVAEAKSPBEKAKSPAIVKEEAKSP	720
OY	289	RAKSGVGEGLATPRAKKENDAKSSOSVGELTLPBLSLKPEKKVAEABKKEVBEAKKAKDQ	348
Dd	721	KAKSPVBEAAKTPBEAKSPVKEEAKSP-----PEKAKSPB--KAKTIDVKSPEAKTPAKEE	773
OY	349	KEEDRRNYPTITYKTTLEIAESD-----VEVKKAELEL--VKEEAKSPNEE--	394
Dd	774	ARSPADKPEKAKSPAIVKEEVKSPBEKAKSPLKADAKPAPEKI.PKKEEVKSPVKEEAKSPQEV	833
OY	395	KVKQAKAEVSKKAEATRLKIKITDRKKAEBEAKRAAEEDKVEK-----	440
Dd	834	KVKPEPKKAAEBEKA.PATPKTEBKDSK--BEAPKKEAPKKEVEEKEPAVEKPEKSYVE	891
OY	441	-----PAEQOPAPAPAKAEK-----APAPKPNPAAQOPKAE	472
Dd	892	AKKEAEADKKKVPPEKAPAKAVEKEDAKPEKTEVAKKEPDDAKAKAPESKPAKKKEA-	950
OY	473	KPADQQAEEDYARSSSEBYNRLTQQQPPKTKPA-----QPSTP	511
Dd	951	APKKTJKEEAKKPEEK-----PKTEAKKEDDKTJLSKPSKP	989

RESULT 8
T09127
probable erythrocyte-binding protein MAEBL - Plasmodium yoelii
C:Species: Plasmodium yoelii
C:Date: 11-Jun-1999 #sequence #revision 11-Jun-1999 #text_change 11-Jun-1999
C:Accession: T09127
R:Kappe, S.H.I.; Noe, A.R.; Fraser, T.S.; Blair, P.L.; Adams, J.H.
Proc. Natl. Acad. Sci. U.S.A. 95, 1230-1235, 1998
A:Title: A family of chemicic erythrocyte binding proteins of malaria parasites
A:Reference number: Z16577; MIMD:98115903; PMID:9448314
A:Accession: T09127
A:Status: Preliminary; translated from GB/EMBL/DDbJ
A:Molecule type: DNA
A:Residues: 1-1701 <KAP>
A:Cross-references: EMBL:AF031886; NID:q2947227; PID:q2947228
A:Experimental source: subspores yoelii; strain YM

A:introns: 62/1; 1648/1; 1674/2; 1697/1
 C:Keywords: alternative splicing; cell binding; erythrocyte invasion
 Query Match 14.2%; Score 361; DB 2; Length 1701;
 Best Local Similarity 29.18%; Pred. No. 6.2e-07;
 Matches 136; Conservative 76; Mismatches 171; Indels 84; Gaps 17

```

Oy  56 QAEQGEQPKUDSRDAR--KEVEYVKTI VGSYAKSTKKRHITVALVNELNI--- 110
    ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db  1107 EAQGAAMAKAAEEERRKAAEAVKAAEEERRKIEAKKAAEEERRKIEAEKAAEEERRRIEAE 1166
    ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Oy  111 -KNEYLWKIYESTSESQULIMESRSKVDEAVSGFKEDSSSSSDSTKPEASDTA-K 168
    ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db  1167 KKAEEERRKIEAAKAAEEERRRIEAKKAAEEERRKIE--AKKAAEEERRKAAEAVKAAE 1222
    ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Oy  169 PNKTEPGEKVAEAKKVVEAEKKAADQDE-----DRANPYITVKTLELEIAESDV 220
    ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db  1224 AKKAAEAAKAAEEERRKAAEAAKKALERKKSEAAKKALERKKAAEAAKAAEKKAAE-- 128
    ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

Oy 221 EVKKAELIVKANEPRDEQIK-----QAEAEVESKQAEATRLKIKITDRERAE 272
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 1282 -AKKAEERKKAELAKKAEERKRLQKRPKRRKRRSGKESRRKRRKRLQKQK-KK 1339
Oy 273 EAKRPADAKEQGKPGRAKRGVGLATPDKKENDAKSSDSSVGEETLPSPLKEKKVA 332
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 1340 KKKRLKRRQKQKRRKRLKRR--QKQRRRKKAEAEKAE-----ERRRI 1383
Oy 333 EAEKVEEKKKAEEDQK--EEDRRVYPTNTYKTLLEIAESVVEKKAELIVKEAKEP 390
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 1384 EAEKAEERKKAELAKKAEER-----KRIAE-----KKAEEERKRIEAKK 1427
Oy 391 RNEEKVK--QAKAEVESKKAATRLKIKITDRKKAEEAKRKAEDDKKEKPAEQOP 447
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 1428 AEEERKRIEAEKKAEEERKRIEAV-----KKAEEERKRIEAE--KKAEEERKRI 1474
Oy 448 APAPAEKRPAPAKPENPAEQPK--AEKPADQAEEDYARRSEEE 490
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 1475 EAVKKAEEERKRIEAEKKAEEERKRIEAKKAEERKRIEAEVAKKEE 1521

```

RESULT 9

```

FC50AG
IGA Fc receptor precursor - Streptococcus agalactiae
N/Alternate names: beta antigen
C/Species: Streptococcus agalactiae
C/Date: 30-Jun-1992 #sequence revision 30-Jun-1992 #text_change 16-Jul-1999
C/Accession: S15330; S20240; S17038
R/Jerlstrom, P.G.; Chhatwal, G.S.; Timmls, K.N.
Mol. Microbiol. 5, 843-849, 1991
A/Title: The IGA-binding beta antigen of the c protein complex of Group B streptococci:
A/Reference number: S15330; MUID:91312121; PMID:1857207
A/Accession: S15330
A/Molecule type: DNA
A/Residues: 1-1164 <JER1>
A/Cross-references: EMBL:X59771
A/Accession: S20240
A/Molecule type: protein
A/Residues: 38-48 <JER2>
R/Jerlstrom, P.G.
submitted to the EMBL Data Library, August 1991
A/Reference number: S17038
A/Accession: S17038
A/Molecule type: DNA
A/Residues: 1-914, 'E', 916-1164 <JBS>
A/Cross-references: EMBL:X59771; NID:946522; PIDN:CAA42442.1; PID:946523
C/Superfamily: IGA Fc receptor
C/Keywords: cell wall; immunoglobulin receptor; tandem repeat; transmembrane protein
F.1-37/Domain: signal sequence #status predicted <SIG>
F.38-1164/Product: IGA Fc receptor #status experimental <MAT>
F.199-438/Domain: IGA binding #status predicted <IGA1>
F.433-826/Domain: IGA binding #status predicted <IGA2>
F.827-945/Region: proline-rich repeats
F.946-1131/Domain: cell wall-spanning #status predicted <CMS>
F.1132-1159/Domain: transmembrane #status predicted <TM>

```

```

Query Match 13.9%; Score 351.5; DB 1; Length 1164;
Best Local Similarity 28.2%; Pred. No. 1e-06;
Matches 143; Conservative 82; Mismatches 177; Indels 105; Gaps 21;

Oy 1 MFASKSRKHYKSIKRFKSVG-ASVVVAISLVNGSVVHAHTENGAQOVPTSSRA-----NES 55
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 1 MFKSNYKRYKSIKRFKSVGVAASVAVASLFWGSAVHAHSELVKDDSVKTEVAAPPSMA 60
Oy 56 QAEQGEOPKLDSEBDKAR-----KEVEEYKKIVGESYAKSTKKHTITVALVNLNN 109
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 61 QTDGNNSSSELETTTMEIRPTTIKKAVEPEVETAGETATDQGEK---QIQCKKNN 117
Oy 110 IKNEYLKYVESISEQLIIMESRSKVDKAVSKFEKDDSSSSSSSSSTKPEASDTAKP 169
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 118 LKNDVNTIILSHQKN-----EFKTKIDE-----TNDSDALLELE----- 152

```

```

Oy 170 NKPTPEGKV-AEAKKVEEAKKAKADQKEEDRRNYPTITYKTLLEIAESDVEV--KKAE 226
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 153 NQFNETNRLHITKQHEVEKXKAKQK-----TLKQSDTFVDSLNDIKELNHQSKQ 204
Oy 227 LELVYKANEPRDEQIKQAEAEVESKQAEATRLKIKITDRERAEAKRPADA-----K 281
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 205 VE-----KMAEQKQITNEDKQSMKTKIEDIRKQA-QQADKKEDAEVKRE 248
Oy 282 EQGPKRAKRGVGLATPDKKENDAKSSDSSVGEETLPSPLKEKKVAEAEKVEA 341
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 249 ELGLFSTKAGLQOEIOEHVKETSSSEENTQKDEHYANSLQNLQKSLLELDKAT-- 305
Oy 342 KKAEDQKEEDRRNYPTNTYKTLLE--IAESDVEVKA--EDELVEAEAKEPNBE-- 394
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 306 --TNEQATQYKQNFLENQKLNKEIQPLIKETNTVKLTKAMSESLQYVEKEK--HNSLAN 360
Oy 395 ----KVKQAKAEVESKKAATRLKIKITDRKKAEEBAKRR--AAEDRVKKEKPAEQP 445
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 361 LEDLVAKSKETIVREYEGKLNQSKLPEL---KQLEEEAHSKLQKVEDFPKFKTSQV 416
Oy 446 QPAPAEKRPAPAKPENPAEQPKAE 472
    |||  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db 417 TPKKRVRRDLAA-----NENNOQKIE 437

```

RESULT 10

```

A60234
IGA Fc receptor precursor - Streptococcus agalactiae (strain SB35)
N/Alternate names: IGA-binding protein; protein Bac
N/Contains: beta antigen
C/Species: Streptococcus agalactiae
C/Date: 08-Dec-1992 #sequence revision 08-Dec-1992 #text_change 26-Aug-1999
C/Accession: A60234; S14595; A60230
R/Heden, L.O.; Fritzb, E.; Lindahl, G.
Eur. J. Immunol. 21, 1481-1490, 1991
A/Title: Molecular characterization of an IGA receptor from group B streptococci: sequen
ents with IGA-binding capacity.
A/Reference number: A60234; MUID:91257158; PMID:2044657
A/Accession: A60234
A/Molecule type: DNA
A/Residues: 1-1134 <HED>
A/Cross-references: EMBL:X58470; NID:946520; PIDN:CAA41384.1; PID:946521
A/Note: The source is designated as group B streptococcus agalactiae
R/Lindahl, G.; Akerstrom, B.; Vaerman, J.P.; Stenberg, L.
Eur. J. Immunol. 20, 2241-2247, 1990
A/Title: Characterization of an IGA receptor from group B streptococci: specificity for
fragments.
A/Reference number: S14595
A/Accession: S14595
A/Molecule type: protein
A/Residues: 'X', 39-48, 'X', 50-52, 'X', 54-56 <LIN>
A/Cross-references: EMBL:X58470; NID:946520; PIDN:CAA41384.1; PID:946521
A/Note: The source is designated as Streptococcus agalactiae
R/Lindahl, G.; Akerstrom, B.; Vaerman, J.P.; Stenberg, L.
Eur. J. Immunol. 20, 2241-2247, 1990
A/Title: Characterization of an IGA receptor from group B streptococci: specificity for
fragments.
A/Reference number: A60230; MUID:91055597; PMID:2242758
A/Accession: A60230
A/Molecule type: protein
A/Residues: 'X', 39-48, 'X', 50-52, 'X', 54-56 <LIN>
C/Superfamily: IGA Fc receptor
C/Keywords: cell wall; immunoglobulin receptor; tandem repeat; transmembrane protein
F.1-37/Domain: signal sequence #status predicted <SIG>
F.38-1134/Product: IGA Fc receptor #status experimental <MAT>
F.199-438/Domain: IGA binding #status predicted <IGA1>
F.439-826/Domain: IGA binding #status predicted <IGA2>
F.827-915/Region: proline-rich repeats
F.916-1101/Domain: cell wall-spanning #status predicted <CMS>
F.1102-1129/Domain: transmembrane #status predicted <TM>

```

```

Query Match 13.7%; Score 346.5; DB 2; Length 1134;
Best Local Similarity 28.0%; Pred. No. 1.5e-06;
Matches 142; Conservative 82; Mismatches 178; Indels 105; Gaps 21;

```

```
Qy      1  HFASSEKXHYSTRPESVG-ASVVVAALWGSVHALENEGATOPYTSSRA-----NES 55
Db      1  MFKSXYEKMKYSTRKFSVGVASVAVRSLFMGSAHAEVLAKDSDVKTEHVAAPRYPSMA 60
Qy      56  QAOGEQCKULDSERDKAR-----KEVEEYVKKI VGESYAKSTKRRTITIVAI VNELANN 109
Db      61  QTDOGNNSSSELETTKMEIPTTDIKKAVERPEKTAGETSAITDGKREK---QLAQOMKN 117
Qy      110 IKNEYLANKIVESTSESOLILMESGRKYDEAVSKFEKSSSSGSDSSTKPREADTAKP 169
Db      118 LKNDVDNTLISHEQN-----BEKTKIDE-----TWSDPALILE-----152
Qy      170 NKPIEPGGKV-AEAKKUYEAKKADQOEEDRRNPITYTYLLELIEAESVEV--KAE 226
Db      153 NQFNETNLHLHKHEHEVEDKDKAKQOK-----TLKOSDTTKZLSINIKELNHQSQ 204
Qy      227 LEIYKVKANBERDEKIKQAEEVESKOAEATRLKKITDPDEBAEBEAKRADAAAA-K 281
Db      205 VE-----KWAEQKGI TNE DKOSM KKI EDIRKOA-QQADKKEDEEVKYRE 248
Qy      282 EOGPKGRAGRKGVPCELATPPDKENDAKSSDSVGEERTLPSPSLKPEKKVAEAKKVEEA 341
Db      249 ELGLPFSTKGLDOEI OEHVKKETSSSEENTQKDEHANSLOMLAQSLELDKAT--- 305
Qy      342 KKKAEDQOEEDRRNPITYTYLLELE--IAESDEVKKA---ELELYKEAKEPRNEB-- 394
Db      306 ---TNEQA TVYKNGFL ENAQKLKEIQPLIKTNVYKLYAWMSBSLEJQVEXELK--HNNSAN 360
Qy      395 -----KYQA KA EVESKKA EATRLEKITDRKKAEEBAKRR---AAEDRYKKEKPAEQ 445
Db      361 LEDIVAASKELIVREYEGKLQSKNLPPL----KOLEEBAHSKLQOVADFPKKFKTSPOV 416
Qy      446 QPAPAPKAEKAPAPKPENPAEQKAE 472
Db      417 TPKKRVKRDIAA-----NENNOQKTE 437
```

[illegible]

```

QY 223 KQAELELVYKXKANEPRDECKIKQAAAEVSK---QAEATRLAKTIKT-D-REAEAEARRA 278
Db 259 KSPEKAKSPEKAKSPVYEVKSPAEEKSPBEKAKSPVKEEKASPEKAKSP 318
QY 279 DAEQOQPKGRARGVGEIATPDCKENDAKSSOSVGEETLPSPSLKPEKVAEAKV 338
Db 319 PVYEEAKSPBEKAKSPVYEEAKSPBEKAKSPVYEEAKS--PEAKSP-VKEAKSPBEKAKSP 375
QY 339 EBAKKA-EQOEEDBRNYPNTYKTLLEIAESDVEYKAELEVKEAKP-----R 391
Db 376 EKAKSPVKEAKSPBEKAKSPV-----EAKSPBEKAKSP-----VKEAKSPKETPKK 424
QY 392 NEKVQO-----AKAEVSKAEATRLKIK-----TRKKAAEEAKR 429
Db 425 EEVKVEPEKPEKVEETAPAPPKVEKSKDEAPKKEAPPAVEKESJAEKADDEADKK 484
QY 430 KAA-----EEDYKVEK---PAEDPOPAPAKAEKAPAKPENPAEOPAKAEKPAQQA 479
Db 485 KAAPAKMEGKEAKPEKTEKTEVAKKEPEADAKKESKPT-EKEPEKPKKEETPAAPVKEKA 543
QY 480 EEDYARSEE-----EYNRILQOOP--PKTEKPAOPST 510
Db 544 KEE-ARKPEKPKTEAKAKEDDKALSKSPKPKTEKAKSSS 584

```

RESULT 12
A44937
kinetoplast-associated protein - Trypanosoma cruzi
N:Alternate names: Probable structural protein KAP
C:Species: Trypanosoma cruzi
C:Date: 28-Apr-1993 #sequence_revision 02-Dec-1994 #text_change 16-Jul-1999
C:Accession: A44937; S27855
R:Gonzalez, A.; Rosales, J.L.; Ley, V.; Diaz, C.
Mol. Biochem. Parasitol. 40, 233-244, 1990
A:Title: Cloning and characterization of a gene coding for a protein (KAP) associated with
A:Reference number: A44937; MUID:90301144; PMID:1694571
A:Accession: A44937
A:Molecule type: DNA
A:Residues: 1-1052 <CON>
A:Cross-references: EMBL:M25364; NID:G162141; PIDN:AAA30209.1; PID:G162142
C:Comment: This protein was detected only in kinetoplasts of replicative stages of the parasite.
C:Comment: This protein is translocated to the mitochondrion without the loss of an amino
C:Superfamily: kinetoplast-associated protein
C:Keywords: duplication; mitochondrion; tandem repeat
I:424-563,582-705,724-862/Region: 9-residue repeats (A-A-R-K-Q-A-E-E)

```

Query Match 13.3% Score 337.5 DB 1 Length 1052.
Best Local Similarity 25.9% Pred. No. 3.2e-06;
Matches 121; Conservative 82; Mismatches 164; Indels 101; Gaps 15

QY 38 ENEGATQVTTSSNNRANESQAEGEQPKLLDSEBDKARKEVEEYKVIQGESYAKSTKRRH 97
Db 493 EEEAARK--QAEEBAARKQAE--BEAARKQAEBAARKRAEEBAARKRAEBAARKRAEE 548

QY 98 TIVVALVNLNNIKYELINKIVYSESQDOLIMNESRSKYDEAVSFEEDDSSSSSDS 157
Db 549 A-----ERKAAEEBAARKRARQARERAEIVKORAAEEBAARKQAEBAARKQAE 601

QY 158 STKPEASDPAKNKNTPEGEKVAEAKKVEEAKKAKQOKEEDRRNYPITYKLTLELEIAE 217
Db 602 AARKQAEBAARKQAE-----BAARKQAEBAARKQAEBAAR-----KQAE 644

QY 218 SDVEVKAELELVKYANEPDECKITQAEAEVESKQAEATRLKLI--TDREBAEEBAK 275
Db 645 EEAARKQAEEEBAARKQAE--EAARKRAEEBAARKAAEAARKQAEAEERKRAEEBA 701

QY 276 R---RAADAEOQPKGRARQGVPELATPDKEANDAKSSSVGEFTLPSPSLKPEKVA 332
Db 702 EKARQAREKAEIVKORRA-----KEBAAK-----QAEEBA 736

QY 333 EAEKVEAEAKKAEDQKEDRRNYPNTYKLTLELEIASDVEVKAELELVKEEAEPN 392

```



```

Db      737  RKQAEAEAAKQAEAEAAKQ-----AEEBAARKQAEBAARKQAE-----  778
Oy      333  EEKVKQAAAEVESKKAETRLKLT--KTDKKAAEEBAARKAAEEDKVKKEPQEQAPAA  450
Db      779  EAARKQAEAEAAKKAEEAAARKCAEEERKKAEEBAARKQAEBAAR--KQAE-----  830
Oy      451  PKAEQAPAPKPEENPAEQPKAEKPPADQQAEEYARSEEEVRLTQQQ  498
Db      831  -----EEAARKQAEEDBAKQAEEDBAKQAEEDBAKTRAE  866

RESULT 13
T34418
hypochemical protein Fl2F3.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34418
R:Fulton, B., Mohlmann, P.
submitted to the EMBL Data Library, July 1998
A:Description: The sequence of C. elegans cosmid Fl1F3.
A:Reference number: Z21521
A:Accession: T34418
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3488 <FUL>
A:Cross-references: EMBL:U80022; PIDN:AAC25885.1; GSPDB:GN00023; CESP:Fl2F3.3
A:Experimental source: strain Bristol N2; clone Fl2F3
C:Genetics:
A:Gene: CESP:Fl2F3.3
A:Map position: 5
A:Introns: 281/3; 332/1; 562/3; 600/3; 1866/3; 1944/3; 3393/1

```

Query Match	13.0%	Score 329.5	DB 2:	Length 3468;
Best Local Similarity	26.8%	Pred. No. 1.9e-05;		
Matches 158;	Conservative	93;	Mismatches 199;	Indels 139;
			Gaps	28;
Qy	3	ASKSERKRVHVAI-----RKFVSQASV--VASLVMGSVHATENEGATOV-----PTSSNRANE	54	
Db	899	AAKKKKVSESVKKGEEKKTKTAGEITVQVESPTSKTITDIDVATETAPETPKKTIK	958	
Qy	55	SOAEQGE-----QPKLDSERDKARKEVEYVKITGESYA-----KSTKKRHITVALVN	105	
Db	959	KKTEKSDSISIQKATDSEKYSKQKODEPTPAVSEOTWTEADSKSKKET-----	1011	
Qy	106	ELNNIKNNYLKNTVESTESQOLIMMSRSKVD--EAVSKFEKSSSSS-----	153	
Db	1012	-----DEKLTDALIAKTKQES--ADEKSLTDQEKITKYSBEDDARKEKELNDKLT	1062	
Qy	154	SSDSSTKEADSTLAKPNKPTPEGKEVAEAKKKEVEAKKAKQKQEDRRANYPITTYKTLEL	213	
Db	1063	ESELATKKAASADTKLT-----BEGQAQA--KKAAVEAAKQKQKXEDQ-----LKLDIT	1107	
Qy	214	FIASDVEVKAQAELELVKVKANEPPEDEQKITQAEAVESKQAEATRLKKITDREEA--	270	
Db	1108	EAAKKAQAAEKTLE-----KQAOIKKAAQADAVKQKQELDEKNNLTAAKKSAAGK	1158	
Qy	271	---EE--AKRRADAKEQGKPRGAKRQVPELAIPTDKENDDASSDSVGEETLPSLSL	325	
Db	1159	LKIEESHAASKQVTEBOAKLDAQK-AKTAEKOT--KLEBDEKSTKESSEKETVDE--	1212	
Qy	326	KPEKKV-----AAEAKVEEA--KKKAEDQKEED--RKNYPNTYKTLEL	366	
Db	1213	KPKKKVLKKTTEKSDSISIQKSETSKTVESAGPESSTQVADARQKQETDEKQKLEA	1272	
Qy	367	FIASDVEVKAQAELELVKEAKEPANNEKVKQAAVENSKKAQATRLKXITDRKGAEE	426	
Db	1273	EIT-----AKKSADSKSLTEA-----BSKLTKA--AEVAARKQKQEKQOELTLDEASKK	1321	
Qy	427	AK-----RKAABEDV--KEKAEQOPAPAPAKAKPKPAPKPEAPQAPK--	471	
Db	1322	AAAEKLEKQSHIKKAAEVDAVAKQKQELEKQRLSESAIIRKADAEKTLKLEBQKKRAE	1381	
Qy	472	-----EKPADQAE-EDYARSRSEENRLLTQOQPTTEKPADS	509	

Db 1382 IALITQKEQETLAQEQSRLEDEAKKSAEKQKLESETKSKQTEAPKES 1430

RESULT 14

S51364

sperm tail-specific protein mst101(2) - fruit fly (*Drosophila hydei*)

C:Species: *Drosophila hydei*

C:Date: 19-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 17-Mar-1999

C:Accession: S51364; S34154

R:Neesen, U.; Padmanabhan, S.; Buennemann, H.
Eur. J. Biochem. 225, 1089-1095, 1994

A>Title: Tandemly arranged repeats of a novel highly charged 16-amino-acid motif represent alpha-helical rods within the extremely elongated spermatzoa of *Drosophila hydei*.

A:Reference number: S51364; MUID:95045538; PMID:7957199

A:Accession: S51364

A:Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-1390 <NEW>

A:Cross-references: EMBL:X73481

R:Neesen, U.; Heinlein, U.A.O.; Buennemann, H.
submitted to the EMBL Data Library, June 1993

A:Reference number: S34153

A:Accession: S34154

A:Molecule type: DNA

A:Residues: 1-163, 'E', 164-236, 'Q', 237-254, 257-320, 'E', 321-1390 <NEW>

A:Cross-references: EMBL:X73481; NID:g313201; PID:g313202

C:Genetics:

A:Gene: mst101(2)

A:Cross-references: FlyBase:FBgn0011816

```

OY 38 ENEGATQVPTSNPANE-SOAEQOGQPKLUBSER---PKAREVEVEYAKKIVGESYAKST 96
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 362 EKAACEKELAKKKGADEKKCEBEAANKKAAIEKKKCEYAKERKEAIEKKCEELAKE 421
OY 94 KGRHTITVALVNEINNIKNBYLNKIVESTSESOTOLIMNESHKVDYEAUSKEKSSSS 153
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 422 KE-----AAERKKCEELANNIKKAAIEKK--KCEAAKKKEKMAERKKCEELAKK 468
OY 154 SSDSTPBEADITAKPNKPTEPGEVNAEAKKKVVEAAKKAADOKEDDRNYPITTYLTLE 212
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 469 IKKAAIEKKKCEETAKKGEVAERKKCEELAKKIKYAEIKKKCKKLAKKEKETAEKKCKEK 528
OY 213 -----LEIAESD-----VEYVKAELTELKVANBPRDOXIKOA-----EAEV 250
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 529 AAKRKEAAIEKKKCEKAKKRRKEAIEKKKCEKSAKGR--EAAIEKKKCEKAKKERKEAAE 586
OY 251 ESQOAEATRLKIKITDREAEIEEAKRRADAEQGRPKGRAK-----RGVGEIATPDK 303
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 587 KKKCEBAKAKKEKEVAERKKCEELAKKIKKAAIEKKKCEAAKKEKEAEREKCGELAKKIK 646
OY 304 KENDAKSSDSVGBRELTPSPSLKPE-----KKVAIEKKVEBA--KKRAEDOK 349
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 647 KAAEKKCKCKLAKKEKETAEKKCKCEKAKKRRKEAIEKKKCEAAKKEKEAIEKKCEAA 706
OY 350 EEDERNYPNTYKLTLEIAESDVEVKKAELELVKEAEKPEPNEBEKVQAKAEVESKAE 409
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 707 KKEKEAERKKCEELAKKIK-----KAAIEKKKCKKLAKKKKAGBKNLKKGNKKGKKA- 759
OY 410 ATRLEKIKITDRKKAEIEEAKRRKAAEEDVKYKEKPAEOPQAPAPAKAPAPENPNAEOP 469
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 760 -----LKEKKCKRELAKKKAIEKKCKCEKAEKAEKKKCEYAKKREE--AEKK 810
OY 470 KAEKPADQOAEEDYARSESE-ENYRNLTOOQPPKTEKPAO 507
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 811 KCERTAKKREAAIEKKKCEKAAKKRRKEAEIEKKKCEKTAK 849

```

129757

protein UNC-89 - *Caenorhabditis elegans*

c/species: caenobacillus elegans
C/Date: 15-Oct-1999 #sequence revision 15-Oct-1999 #text change 03-Dec-1999

C;Accession: T29757

R; Du, Z.; Le, T.T.; Wilson, R. submitted to the EMBL Data Lib

A;Description: The sequence of *C. elegans* co-

A:Reference number: Z20679

A;Accession: J29/5/
A;Status: preliminary: tra

A;Molecule type: DN

A;Residues: 1-6642

A:Experimental source: strain Bristol N2; clone C09
A;CROSS-references: EMBL:AF003131; PIDN:AAB54132.1;

C;Genetics:

A;Gene: CESP:unc-89

A: map position: 1
A: introns: 17/2: 108/3: 154/2: 211/2: 265/3: 326/2: 352/3: 426/2: 454/1: 500/1

/3; 5917/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1

Quarter Match

Best Local Similarity

Matches 133;

00 37 TENEGATOVPTSSNPANESQAEQ-CEQ- - - - PKYIDSEBDRAPKVEFEVVKYVGSYA 90

100

Db 1431

91 α

2

Db 1485

151 ov

73

Db 1539

2020v

2

Db 1599

Ov 262

1

Db 1650

QY 307

2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
84

DB 1.709

QY 367

1755

1/65
DB

QY 424

1030
D

1920 22

QY 469

Db 1877

Search completed: November 21, 2003, 13:34:25
Job time : 22.7062 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:22:21 ; Search time 19.5369 Seconds
(without alignments)
1462.401 Million cell updates/sec

Title: US-09-298-523C-13_COPY_263_442
Perfect score: 889
Sequence: 1 IKTDREAREBEAKRRADAKE.....AEEAKRKAABEDKVEKPA 180

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	889	100.0	406	20	AAV49225
2	889	100.0	406	20	AAV32098
3	889	100.0	428	20	AAV49238
4	889	100.0	428	20	AAV32110
5	889	100.0	446	20	AAV49143
6	889	100.0	453	19	AAW55081
7	889	100.0	453	23	ABP54575
8	889	100.0	460	20	AAV49250
9	889	100.0	655	20	AAV49226

10	889	100.0	655	20	AAV32099	Choline binding pr
11	889	100.0	693	23	AAU76890	Pneumococcal surfa
12	889	100.0	693	24	ABU02720	S. pneumoniae type
13	889	100.0	694	21	AAV81653	Streptococcus pneu
14	886	99.7	460	20	AAV32189	N-terminal choline
15	884	99.4	446	20	AAV49140	Amino acid sequenc
16	884	99.4	459	20	AAV49251	N-terminal region
17	884	99.4	459	20	AAV32190	N-terminal choline
18	803.5	90.4	376	20	AAV49231	N-terminal truncat
19	803.5	90.4	376	20	AAV32103	Choline binding pr
20	803.5	90.4	413	20	AAV49247	N-terminal region
21	803.5	90.4	413	20	AAV32186	N-terminal choline
22	803.5	90.4	414	20	AAV49150	Amino acid sequenc
23	803.5	90.4	564	19	AAW62654	C3 binding protein
24	803.5	90.4	581	20	AAV43394	S. pneumoniae SpSa
25	803.5	90.4	631	18	AAW6444	Cbpa, a choline bi
26	803.5	90.4	663	20	AAV49232	Cbpa of serotype R
27	803.5	90.4	663	20	AAV32104	Choline binding pr
28	803.5	90.4	663	20	AAV43395	S. pneumoniae chol
29	803.5	90.4	663	23	AAE29201	Streptococcus pneu
30	803.5	90.4	663	23	ABG60639	Partial sequence o
31	803.5	90.4	663	24	ABP55319	Streptococcus pneu
32	799.5	89.9	701	23	AAU76888	Pneumococcal surfa
33	795	89.4	439	20	AAV49242	N-terminal region
34	795	89.4	439	20	AAV32181	N-terminal choline
35	793	89.2	419	20	AAV49243	N-terminal region
36	793	89.2	419	20	AAV32182	N-terminal choline
37	792	89.1	406	20	AAV49152	Amino acid sequenc
38	792	89.1	425	20	AAV49145	Amino acid sequenc
39	791.5	89.0	414	20	AAV49144	Amino acid sequenc
40	784	88.2	438	20	AAV49141	Amino acid sequenc
41	784	88.2	439	20	AAV49249	N-terminal region
42	784	88.2	439	20	AAV32188	N-terminal choline
43	775.5	87.2	412	20	AAV49151	Amino acid sequenc
44	775.5	87.2	425	20	AAV49248	N-terminal region
45	775.5	87.2	425	20	AAV32187	N-terminal choline

ALIGNMENTS

RESULT 1
ID AAV49225 standard; protein: 406 AA.
AAV49225:
XX
XX AAV49225:
AC
XX
XX
XX 07-FEB-2000 (first entry)
DT
XX
XX
DE Polypeptide R2 of the N-terminal truncate of Cbpa serotype type 4.
XX
XX Choline binding protein A; Cbpa; truncate; immune response; infection;
KM pneumococcal bacterium; vaccine.
XX
XX
OS Streptococcus pneumoniae.
XX
XX
XX WO951187.A2.
PN
XX
XX
XX 14-OCT-1999.
PD
XX
XX
XX 07-APR-1999; 99WO-US07668.
PF
XX
XX 07-APR-1998; 98US-0056019.
PR
XX 07-APR-1998; 98US-0080878.
PR
XX
XX (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
PA
XX
XX Tuomanen EI, Masure HR;
PI
XX
XX WPI, 1999-620161/53.
DR N-PSDB; AA231401.
XX
XX Novel polypeptides, used to develop products for the diagnosis,
PT

PT prevention and treatment of pneumococcal infections -
XX
XX Claim 2; Page 7-8; 85pp; English.
CC The invention provides novel isolated polypeptides comprising the amino
CC acid sequence of an N-terminal choline binding protein A (Cbpa) truncate
CC (CbpaT). The polypeptides can be selected from sequences shown in
CC AA49225, AA49227, AA49230, AA49231, AA49233 and AA49238. The
CC polypeptides can be used for inducing an immune response in a subject
CC which has been exposed to or infected with a pneumococcal bacterium. They
CC can also be used for preventing infection by a pneumococcal bacterium.
CC Vaccines comprising the polypeptides or encoding nucleic acids can be
CC used for treating a subject infected with or exposed to a pneumococcal
CC bacterium. Antibodies specifically binding the polypeptides can be used
CC for detection and diagnosis and for preventing pneumococcal attachment to
CC a mucosal surface. The products can be used in humans and other animals
CC such as domestic animals, such as feline or canine subjects, farm animals
CC animals (whether in the wild or in a zoological garden), research
CC animals, such as mice, rats, rabbits, goats, sheep, pigs, dogs, cats,
CC i.e. for veterinary medical use.
XX
XX Sequence 406 AA;
SQ
Query Match 100.0%; Score 889; DB 20; Length 406;
Best Local Similarity 100.0%; Pred. No. 1.2e-62;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IKTDRBEAEERAKRRADAKGKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPS 60
DB 227 IKTDRBEAEERAKRRADAKGKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPS 286
QY 61 PSLRPEKKVAEAEKKVEBAKKKADQKEEDRRNPTNTYKLTLEIASDVYKKAEL 120
DB 287 PSLRPEKKVAEAEKKVEBAKKKADQKEEDRRNPTNTYKLTLEIASDVYKKAEL 346
QY 121 VKEAKEPRNEBKVKQAKAEVSKKAEATRLKIKTDKKKAEBAKRAAEEDVKYKCPA 180
DB 347 VKEAKEPRNEBKVKQAKAEVSKKAEATRLKIKTDKKKAEBAKRAAEEDVKYKCPA 406
RESULT 2
AA492098
ID AA492098 standard; Protein; 406 AA.
AC AA492098;
XX
XX 01-FEB-2000 (first entry)
DT
XX
XX Choline binding protein A (Cbpa) truncate R2.
DE
XX Choline binding protein; Cbpa; truncate; adhesin; immunogen;
KW vaccine; diagnosis; therapy; pneumococcus; sepsis; meningitis;
KW otitis media; pneumonia.
XX
XX Streptococcus pneumoniae.
OS Synthetic.
XX
XX WO9951187-A2.
PN
XX 14-OCT-1999.
PD
XX 07-APR-1999; 99WO-US07669.
PF
XX 07-APR-1998; 98US-0056019.
PR 07-APR-1998; 98US-0080878.
XX
XX (STUD-) ST JUDE CHILDREN'S RES HOSPITAL.
PA (MEDI-) MEDIMUNE INC.
PI Tuomanen EI, Masure HR, Wizemann TM, Johnson LS, Koenig S;
XX MPI; 1999-633690/54.
DR

DR N-PSDB; AA234451.
XX
XX New N-terminal choline binding protein A truncate polypeptides used to
PT develop products for the diagnosis, prevention and treatment of
PT pneumococcal infections -
XX
XX Claim 2; Page 85-88; 160pp; English.
PS
CC The present sequence represents polypeptide R2, a new N-terminal
CC choline binding protein A (Cbpa) truncate of Streptococcus pneumoniae
CC serotype type 4. Host-vector systems for production of R2 and other
CC N-terminal Cbpa truncates are provided. Important features of R2
CC include: (1) complete correlation of bioactivity with full-length
CC Cbpa for recognition of purified glycoconjugate receptor analogues,
CC lung cells and animal models; and (2) cross-protection between type
CC 4 derived agents and bacteria in in vitro assays using other
CC serotypes (e.g. 6B and 2), important for use in vaccine, prophylaxis
CC and therapy. The invention relates generally to novel N-terminal
CC Cbpa truncates and polynucleotides encoding them, host-vector
CC systems, and antibodies that specifically bind to the truncates.
CC The invention also relates to vaccines including N-terminal Cbpa
CC truncates, which provide protection or elicit protective antibodies
CC to bacterial infection, specifically pneumococcus, and to use of
CC antibodies and antagonists against such polypeptides in diagnosis
CC and passive immunotherapy. The polypeptides and/or polynucleotides
CC are also useful as competitive inhibitors of bacterial adhesion of
CC pneumococcus.
XX
XX Sequence 406 AA;
SQ
Query Match 100.0%; Score 889; DB 20; Length 406;
Best Local Similarity 100.0%; Pred. No. 1.2e-62;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IKTDRBEAEERAKRRADAKGKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPS 60
DB 227 IKTDRBEAEERAKRRADAKGKPKGRAGVPGELATPDKKENDAKSSDSSVGEETLPS 286
QY 61 PSLRPEKKVAEAEKKVEBAKKKADQKEEDRRNPTNTYKLTLEIASDVYKKAEL 120
DB 287 PSLRPEKKVAEAEKKVEBAKKKADQKEEDRRNPTNTYKLTLEIASDVYKKAEL 346
QY 121 VKEAKEPRNEBKVKQAKAEVSKKAEATRLKIKTDKKKAEBAKRAAEEDVKYKCPA 180
DB 347 VKEAKEPRNEBKVKQAKAEVSKKAEATRLKIKTDKKKAEBAKRAAEEDVKYKCPA 406
RESULT 3
AA49238
ID AA49238 standard; Protein; 428 AA.
AC AA49238;
XX
XX 07-FEB-2000 (first entry)
DT
XX
XX Choline binding protein A (Cbpa) serotype type 4 amino acid sequence.
DE
XX Choline binding protein A; Cbpa; truncate; immune response; infection;
KW pneumococcal bacterium; vaccine.
KW
XX Streptococcus pneumoniae.
OS
XX
XX WO9951187-A2.
PN
XX 14-OCT-1999.
PD
XX 07-APR-1999; 99WO-US07668.
PF
XX 07-APR-1998; 98US-0056019.
PR 07-APR-1998; 98US-0080878.
XX
XX (STUD-) ST JUDE CHILDREN'S RES HOSPITAL.
PA
XX

PI	Tuomanen EI, Masure HR;
XX	
DR	WPI, 1999-620161/53.
XX	
PT	Novel polypeptides, used to develop products for the diagnosis,
PR	prevention and treatment of pneumococcal infections -
XX	
PS	Claim 7, Page 7; 85pp; English.
XX	
CC	The invention provides novel isolated polypeptides comprising the amino
CC	acid sequence of an N-terminal choline binding protein A (Cbpa) truncate
CC	(CbpaT). The polypeptides can be selected from sequences shown in
CC	AAI99225, AAI99227, AAI99230, AAI99231, AAI99233 and AAI99238. The
CC	polypeptides can be used for inducing an immune response in a subject
CC	which has been exposed to or infected with a pneumococcal bacterium.
CC	It can also be used for preventing infection by a pneumococcal bacterium.
CC	Vaccines comprising the polypeptides or encoding nucleic acids can be
CC	used for treating a subject infected with or exposed to a pneumococcal
CC	bacterium. Antibodies specifically binding the polypeptides can be used
CC	for detection and diagnosis and for preventing pneumococcal attachment to
CC	a mucosal surface. The products can be used in humans and other animals
CC	such as domestic animals, such as feline or canine subjects, farm animals
CC	such as bovine, equine, caprine, ovine, and porcine subjects, wild
CC	animals (whether in the wild or in a zoological garden), research
CC	animals, such as mice, rats, rabbits, goats, sheep, pigs, dogs, cats,
CC	i.e. for veterinary medical use.
XX	
SQ	Sequence 428 AA;
Query Match	100.0%; Score 889; DB 20; Length 428;
Best Local Similarity	100.0%; Pred. No. 1.3e-62;
Matches 180; Conservative	0; Mismatches 0; Indels 0; Gaps 0
QY	1 IKTRDEEEAEAKRRADAEKGGPKGRAGVPGELATPDKKENNAKSDDSSVGSEETLPS 60
DB	227 IKTBREAEAEAKRRADAKEGCKPKGRAGVPGSLATPPDKKENAKSDDSSVGSEETLPS 286
DY	61 PSLRPEKKVAEAEKKVEBAKKKAEDQKEEDRRNYPNTYKTLELIAESDVEVKKADEL 120
DB	287 PSLRPEKKVAEAEKKVEBAKKKAEDQKEEDRRNYPNTYKTLELIAESDVEVKKADEL 346
QY	121 VKEEKERNEBKVOAKAIEVSKKAENTRLFKITDRKKAEKKRAQAEDDKYKEKPA 180
DB	347 VKEEKERNEBKVOAKAIEVSKKAENTRLFKITDRKKAEKKRAQAEDDKYKEKPA 406
RESULT 4	
AAI92110	
ID	AAI92110 standard; Protein; 428 AA.
XX	
AC	AAI92110;
XX	
DT	01-FEB-2000 (first entry)
XX	
DE	Choline binding protein A (Cbpa) N-terminal truncate.
XX	
KM	Choline binding protein; Cbpa; truncate; adhesin; immunogen;
KW	vaccine; diagnosis; therapy; pneumococcus; sepsis; meningitis;
KW	otitis media; pneumonia.
OS	Streptococcus pneumoniae.
XX	
OS	Synthetic.
XX	
PJ	MO9951188-A2.
XX	
PD	14-OCT-1999.
XX	
PF	07-APR-1999; 99WO-USO7669.
XX	
FR	07-APR-1998; 98US-0056019.
PR	07-APR-1998; 98US-0080878.
XX	
SA	(SUUD-) ST JUDE CHILDREN'S RES HOSPITAL.

PA	(MEDI-) MEDIMMUNE INC.
XX	Tuomanen EI, Masure HR, Wizemann TM, Johnson LS, Koenig S;
XX	WPI; 1999-633690/54.
DR	New N-terminal choline binding protein A truncate polypeptides, used to
XX	develop products for the diagnosis, prevention and treatment of
PT	pneumococcal infections -
PS	Claim 7; Page 119-122; 160pp; English.
CC	The present sequence represents a new N-terminal choline binding
CC	protein A (Cbpa) truncate that maintains the tertiary structure of
CC	<i>Streptococcus pneumoniae</i> serotype type 4 Cbpa native protein.
CC	Host-vector systems for production of N-terminal Cbpa truncates are
CC	provided. The invention relates generally to novel N-terminal
CC	Cbpa truncates and polynucleotides encoding them, host-vector
CC	systems, and antibodies that specifically bind to the truncates.
CC	The invention also relates to vaccines including N-terminal Cbpa
CC	truncates, which provide protection or elicit protective antibodies
CC	to bacterial infection, specifically pneumococcus, and to use of
CC	antibodies and antagonists against such polypeptides in diagnosis
CC	and passive immunotherapy. The polypeptides and/or polynucleotides
CC	are also useful as competitive inhibitors of bacterial adhesion of
CC	pneumococcus.
SQ	Sequence 428 AA;
Query Match	100.0%; Score 889; DB 20; Length 428;
Best Local Similarity	100.0%; Pred. No. 1,3e-62;
Matches 180; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	1 IKTRREAEFEBAKRRADAKGPKPGRAKRGVPEGLATPPDKKENDAKSSDVGSETLPS 60
DB	227 IKTREBEFEBAKRRADAKGEQKEKGRAKKGVPEGLATPPDKKENDAKSSDVGSETLPS 286
OY	61 PSLRPEKKVAEAKKVEBAKKKAEDQEKEDBRNYPNTNYKTLELEIAESDVEVKKALEL 120
DB	287 PSLRPEKKVAEAKKVEBAKKKAEDQEKEDBRNYPNTNYKTLELEIASDVVKKALEL 346
OY	121 VKEBAKEPRNEBKTKAKAEVESKKAETRLKETKDRAKKAAEEBAKRKAABEDKYKEKPA 180
DB	347 VKERAKEPRNEBKTKAKAEVESKKAETRLKETKDRAKKAAEEBAKRKAABEDKYKEKPA 406
RESULT 5	
AAAY49143	
ID	AAAY49143 standard; Protein; 446 AA.
AC	AAAY49143;
XX	
DT	17-JAN-2000 (first entry)
DE	Amino acid sequence of choline-binding protein fragment #7.
XX	
KW	Vaccinate surface binding protein; alpha helix; choline binding protein;
KW	tuncated; invasive bacterial infection; otitis media; sepsis;
KW	meningitis; lobar pneumonia infection; antibody; immature immune system;
KW	immunocompromised.
OS	<i>Streptococcus pneumoniae</i> .
PJ	WO9951266-A2.
PD	14-OCT-1999.
PF	06-APR-1999; 99MO-USO7680.
FR	07-APR-1998; 98US-0080878.
EX	15-MAY-1998; 98US-0085743.
PA	(MEDI-) MEDIMMUNE INC.

```
XX
PI Wizemann TM, Koenig S, Johnson LS;
XX
DR WPI; 1999-601465/51.
DR N-PSDB; AA231083.
XX
PT New pneumococcal proteins useful as vaccines and for diagnosis of
PT pneumococcal infections -
XX
PS Claim 10; Page 72-73; 98pp; English.
XX
CC AA49137-Y49152 are amino acid sequences that are fragments of choline
CC binding proteins (CBP). The fragments of the protein are the alpha helix
CC forming parts of the CBPs from Streptococcus pneumoniae. The
CC polypeptides do not contain the actual choline binding fragment. The
CC polypeptides and the nucleotide sequences that encode them
CC (AA31077-231092) are used in the invention, which relates to polypeptide
CC truncates of a pneumococcal surface binding protein containing the highly
CC conserved immunogenic alpha helical portion and no choline binding
CC portion. The polypeptides are used as immunogens in a bacterial vaccine.
CC The vaccine can be used for preventing (immunizing) or treating invasive
CC bacterial (especially pneumococcal) infections, especially otitis media
CC (caused by S.pneumoniae), sepsis, meningitis and lobar pneumonia
CC infections. Antibodies raised against the polypeptide are useful for
CC detection, prevention (passive immunity) and treatment of S. pneumoniae
CC infections. The vaccines are especially useful in immunocompromised
CC patients, those with an immature immune system, or patients with an on
CC going pneumococcal infection. The vaccine avoids unnecessary expense and
CC provides broad protection against a range of pneumococcal serotypes and
CC it produces an improved and enhanced effect in preventing bacterial
CC infections.
XX
SQ Sequence 446 AA;
XX
Query Match 100.0%; Score 889; DB 20; Length 446;
Best Local Similarity 100.0%; Pred. No. 1.3e-62;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 IKTDREBEAEEBKAERADAKGKPKGRAGVPGELATPPDKKENDAKSSDSSVGEETLPS 60
DB 228 IKTDREBEAEEBKAERADAKGKPKGRAGVPGELATPPDKKENDAKSSDSSVGEETLPS 287
OY 61 PSLKPEKKVAEAEKKVEAEKKKADOKEDRRNYPNTYTKTLELEIASDVEVKAELEL 120
DB 268 PSLKPEKKVAEAEKKVEAEKKKADOKEDRRNYPNTYTKTLELEIASDVEVKAELEL 347
OY 121 VKEAKEPRNEBKVKQAKAEVSKKAEATRLKIKTDKKAEEBKRAAEEDVKKEKPA 180
DB 348 VKEAKEPRNEBKVKQAKAEVSKKAEATRLKIKTDKKAEEBKRAAEEDVKKEKPA 407
XX
RESULT 6
AAM55081
ID AAM55081 standard; Protein; 453 AA.
XX
AC AAM55081;
XX
DT 02-OCT-1998 (first entry)
XX
DE Streptococcus pneumoniae SP0023 protein.
XX
KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
KW detection; pneumonia; otitis media; meningitis.
XX
OS Streptococcus pneumoniae.
XX
PN MO9818930-A2.
XX
PD 07-MAY-1998.
XX
PF 30-OCT-1997; 97WO-US19422.
XX
PR 31-OCT-1996; 96US-0029960.
```

```
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Choi GH, Hromockyj A, Johnson LS, Kunsch CA;
XX
DR WPI; 1998-272224/24.
DR N-PSDB; AAV27342.
XX
PT Nucleic acid encoding antigenic peptide(s) from Streptococcus
PT pneumoniae - or their epitope-containing fragments, useful in
XX protective or therapeutic vaccines, and for diagnosis
XX
PS Claim 11; Page 56; 118pp; English.
XX
CC The present sequence represents a protein from Streptococcus pneumoniae.
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein
CC can be useful in vaccines for inducing protective antibodies against
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
CC are used to detect Streptococcus infection (by usual hybridisation or
CC amplification methods), also for isolating Streptococcus genes or their
CC allelic variants. The protein can be used similarly to detect specific
CC antibodies in standard immunoassays, especially for diagnosing or
CC monitoring infections. Antibodies which bind the protein are used to
CC detect corresponding antigens, to purify the protein and for passive
CC immunisation (optionally coupled to a toxin). Vaccines are administered,
CC e.g. by injection, orally or through the skin, typically at 0.01-1000
CC (especially 10-300) mu g/ml per dose.
XX
SQ Sequence 453 AA;
XX
Query Match 100.0%; Score 889; DB 19; Length 453;
Best Local Similarity 100.0%; Pred. No. 1.4e-62;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 IKTDREBEAEEBKAERADAKGKPKGRAGVPGELATPPDKKENDAKSSDSSVGEETLPS 60
DB 25 IKTDREBEAEEBKAERADAKGKPKGRAGVPGELATPPDKKENDAKSSDSSVGEETLPS 84
OY 61 PSLKPEKKVAEAEKKVEAEKKKADOKEDRRNYPNTYTKTLELEIASDVEVKAELEL 120
DB 85 PSLKPEKKVAEAEKKVEAEKKKADOKEDRRNYPNTYTKTLELEIASDVEVKAELEL 144
OY 121 VKEAKEPRNEBKVKQAKAEVSKKAEATRLKIKTDKKAEEBKRAAEEDVKKEKPA 180
DB 145 VKEAKEPRNEBKVKQAKAEVSKKAEATRLKIKTDKKAEEBKRAAEEDVKKEKPA 204
XX
RESULT 7
ABP54575
ID ABP54575 standard; Protein; 453 AA.
XX
AC ABP54575;
XX
DT 04-SEP-2002 (first entry)
XX
DE S. pneumoniae SP023 protein sequence SEQ ID NO:38.
XX
KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
KW antibacterial; Streptococcal infection; detection.
XX
OS Streptococcus pneumoniae.
XX
PN US2002061545-A1.
XX
PD 23-MAY-2002.
XX
PF 22-JAN-2001; 2001US-0765272.
XX
PR 30-OCT-1997; 97US-0961083.
XX
PA (CHOI/) CHOI G H.
PA (KUNS/) KUNSCH C A.
```

PA (BARA/) BARASH S C.
PA (DILL/) DILLON P J.
PA (DOUG/) DOUGHERTY B.
PA (FANN/) FANNON M R.
PA (ROSE/) ROSEN C A.
XX
PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR,
PI Rosen CA;
XX
XX WPI: 2002-479261/51.
DR N-PSDB; ABQ84810.
XX
XX New Streptococcus pneumoniae antigens, useful for detecting
PT Streptococcus and for preventing or attenuating disease caused by
PT Streptococcus infection -
XX
XX Claim 11, Page 24-25; 70pp; English.
XX
XX ABQ84792 to ABQ84904 represents nucleic acids which encode the
CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.
CC The S. pneumoniae antigens have antibacterial activity and can be
CC used in vaccines. The S. pneumoniae antigens can also be used to
CC prevent or attenuate a Streptococcal infection in an animal. The
CC polynucleotides encoding the S. pneumoniae antigens can be used to
CC detect Streptococcus nucleic acids. ABQ84905 to ABQ85130 represent
CC primers used in the cloning of S. pneumoniae ORFs (open reading frames)
CC which are used in an example from the present invention.
XX
XX Sequence 453 AA;
SQ
Query Match 100.0%; Score 889; DB 23; Length 453;
Best Local Similarity 100.0%; Pred. No. 1,4e-62;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IKTDREBAEEBAKRRADAKGKPKGRAGVPGELATPDCKENDAKSSDSVGEETLPS 60
DB 25 IKTDREBAEEBAKRRADAKGKPKGRAGVPGELATPDCKENDAKSSDSVGEETLPS 84
QY 61 PSLKPEKKVAEAEKKVEAKKKAEDOKEDRRNPPTNTYKTLLEIAESDVEYKKAELEL 120
DB 85 PSLKPEKKVAEAEKKVEAKKKAEDOKEDRRNPPTNTYKTLLEIAESDVEYKKAELEL 144
QY 121 VKEBAKEPRNEBEKVOKAKAEVESKKAETRLKIKTDRKKAEEBAKKAEEBDKVKEKPA 180
DB 145 VKEBAKEPRNEBEKVOKAKAEVESKKAETRLKIKTDRKKAEEBAKKAEEBDKVKEKPA 204
RESULT 8
AAV49250
ID AAV49250 standard; Protein; 460 AA.
XX
AC AAV49250;
XX
DT 07-FEB-2000 (first entry)
XX
XX N-terminal region of CbpA polypeptide Ntype4.
XX
XX Choline binding protein A; CbpA; truncate; immune response; infection;
KW pneumococcal bacterium; vaccine.
XX
OS Streptococcus pneumoniae.
XX
PN WO9951187-A2.
XX
PD 14-OCT-1999.
XX
PF 07-APR-1999; 99WO-US07668.
XX
PR 07-APR-1998; 98US-0056019.
PR 07-APR-1998; 98US-0080878.
XX
PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
XX

PI Tuomanen EI, Masure HR;
XX
XX WPI: 1999-620161/53.
DR
XX Novel polypeptides, used to develop products for the diagnosis,
PT prevention and treatment of pneumococcal infections -
XX
XX Disclosure; Fig 2A-B; 85pp; English.
XX
XX The invention provides novel isolated polypeptides comprising the amino
CC acid sequence of an N-terminal choline binding protein A (CbpA) truncate
CC (CbpAT). The polypeptides can be selected from sequences shown in
CC AAV49225, AAV49227, AAV49230, AAV49231, AAV49233 and AAV49238. The
CC polypeptides can be used for inducing an immune response in a subject
CC which has been exposed to or infected with a pneumococcal bacterium. They
CC can also be used for preventing infection by a pneumococcal bacterium.
CC Vaccines comprising the polypeptides or encoding nucleic acids can be
CC used for treating a subject infected with or exposed to a pneumococcal
CC bacterium. Antibodies specifically binding the polypeptides can be used
CC for detection and diagnosis and for preventing pneumococcal attachment to
CC a mucosal surface. The products can be used in humans and other animals
CC such as bovine, equine, caprine, ovine, and porcine subjects, farm animals
CC such as dogs, cats, and rabbits, and in a zoological garden, research
CC animals, such as mice, rats, rabbits, goats, sheep, pigs, dogs, cats,
CC i.e. for veterinary medical use.
XX
XX Sequence 460 AA;
SQ
Query Match 100.0%; Score 889; DB 20; Length 460;
Best Local Similarity 100.0%; Pred. No. 1,4e-62;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IKTDREBAEEBAKRRADAKGKPKGRAGVPGELATPDCKENDAKSSDSVGEETLPS 60
DB 242 IKTDREBAEEBAKRRADAKGKPKGRAGVPGELATPDCKENDAKSSDSVGEETLPS 301
QY 61 PSLKPEKKVAEAEKKVEAKKKAEDOKEDRRNPPTNTYKTLLEIAESDVEYKKAELEL 120
DB 302 PSLKPEKKVAEAEKKVEAKKKAEDOKEDRRNPPTNTYKTLLEIAESDVEYKKAELEL 361
QY 121 VKEBAKEPRNEBEKVOKAKAEVESKKAETRLKIKTDRKKAEEBAKKAEEBDKVKEKPA 180
DB 362 VKEBAKEPRNEBEKVOKAKAEVESKKAETRLKIKTDRKKAEEBAKKAEEBDKVKEKPA 421
RESULT 9
AAV49226
ID AAV49226 standard; Protein; 655 AA.
XX
AC AAV49226;
XX
DT 07-FEB-2000 (first entry)
XX
XX CbpA of serotype 4 amino acid sequence.
XX
XX Choline binding protein A; CbpA; truncate; immune response; infection;
KW pneumococcal bacterium; vaccine.
XX
OS Streptococcus pneumoniae.
XX
PN WO9951187-A2.
XX
PD 14-OCT-1999.
XX
PF 07-APR-1999; 99WO-US07668.
XX
PR 07-APR-1998; 98US-0056019.
PR 07-APR-1998; 98US-0080878.
XX
PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
XX
PI Tuomanen EI, Masure HR;

XX WPI; 1999-620161/53.
DR N-PSDB; AA231402.
XX
PT Novel polypeptides, used to develop products for the diagnosis,
PT prevention and treatment of pneumococcal infections
XX
PS Disclosure; Page 9; 85pp; English.
XX
CC The invention provides novel isolated polypeptides comprising the amino
CC acid sequence of an N-terminal choline binding protein A (Cbpa) truncate
CC (CbpaT). The polypeptides can be selected from sequences shown in
CC AA9225, AA9227, AA9230, AA9231, AA9232 and AA9238. The
CC polypeptides can be used for inducing an immune response in a subject
CC which has been exposed to or infected with a pneumococcal bacterium. They
CC can also be used for preventing infection by a pneumococcal bacterium.
CC Vaccines comprising the polypeptides or encoding nucleic acids can be
CC used for treating a subject infected with or exposed to a pneumococcal
CC bacterium. Antibodies specifically binding the polypeptides can be used
CC for detection and diagnosis and for preventing pneumococcal attachment to
CC a mucosal surface. The products can be used in humans and other animals
CC such as domestic animals, such as feline or canine subjects, farm animals
CC such as bovine, equine, caprine, ovine, and porcine subjects, wild
CC animals (whether in the wild or in a zoological garden), research
CC animals, such as mice, rats, rabbits, goats, sheep, pigs, dogs, cats,
CC i.e. for veterinary medical use.
XX
SQ Sequence 655 AA;
Query Match 100.0%; Score 889; DB 20; Length 655;
Best Local Similarity 100.0%; Pred. No. 2.1e-62;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IKTRBEAEAEKRRADAKGKPKGRKRGVPGELATPPDKKENDAKSSDSSVGEETLPS 60
DB 227 IKTRBEAEAEKRRADAKGKPKGRKRGVPGELATPPDKKENDAKSSDSSVGEETLPS 286
QY 61 PSLRPEKVAEAEKVEAEKKAEDOKEDRRNPTNTYKTLLEIAESDVEVKAELEL 120
DB 287 PSLRPEKVAEAEKVEAEKKAEDOKEDRRNPTNTYKTLLEIAESDVEVKAELEL 346
QY 121 VKEAEKEPRNEEKYKQAKAEVESKKAATRLKIKTRKKAEEBAKRAAEEDVKKEKPA 180
DB 347 VKEAEKEPRNEEKYKQAKAEVESKKAATRLKIKTRKKAEEBAKRAAEEDVKKEKPA 406
RESULT 10
AA92099
ID AA92099 standard; Protein; 655 AA.
XX
AC AA92099;
XX
DT 01-FEB-2000 (first entry)
XX
DE Choline binding protein A (Cbpa).
XX
KW Choline binding protein; Cbpa; adhesin; immunogen;
KW vaccine; diagnosis; therapy; pneumococcus; sepsis; meningitis;
KW otitis media; pneumonia.
XX
OS Streptococcus pneumoniae.
XX
PN MO9951188-A2.
XX
PD 14-OCT-1999.
XX
PF 07-APR-1999; 99WO-US07669.
XX
PR 07-APR-1998; 98US-0056019.
XX
PR 07-APR-1998; 98US-0080878.
XX
PA (SJDUD-) ST JUDE CHILDREN'S RES HOSPITAL.
PA (MEDI-) MEDIMUNE INC.

XX
PI Tuomannen EI, Maure HR, Wizemann TM, Johnson LS, Koenig S;
XX WPI; 1999-633690/54.
DR N-PSDB; AA234452.
XX
PT New N-terminal choline binding protein A truncate polypeptides, used to
PT develop products for the diagnosis, prevention and treatment of
PT pneumococcal infections
XX
PS Disclosure; Page 88-93; 160pp; English.
XX
CC This sequence represents the choline binding protein A (Cbpa)
CC of Streptococcus pneumoniae serotype type 4. The invention provides
CC novel N-terminal Cbpa truncated polypeptides (see AA92098-110 and
CC AA92179-90) and polynucleotides encoding them, host-vector systems,
CC and antibodies that specifically bind to the polypeptides. The
CC invention also relates to vaccines including the polypeptides,
CC which provide protection or elicit protective antibodies to
CC bacterial infection, specifically pneumococcus, and to antibodies
CC and antagonists against such polypeptides for use in diagnosis and
CC passive immunotherapy. The polypeptides and/or polynucleotides are
CC also useful as competitive inhibitors of bacterial adhesion of
CC pneumococcus.
XX
SQ Sequence 655 AA;
Query Match 100.0%; Score 889; DB 20; Length 655;
Best Local Similarity 100.0%; Pred. No. 2.1e-62;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IKTRBEAEAEKRRADAKGKPKGRKRGVPGELATPPDKKENDAKSSDSSVGEETLPS 60
DB 227 IKTRBEAEAEKRRADAKGKPKGRKRGVPGELATPPDKKENDAKSSDSSVGEETLPS 286
QY 61 PSLRPEKVAEAEKVEAEKKAEDOKEDRRNPTNTYKTLLEIAESDVEVKAELEL 120
DB 287 PSLRPEKVAEAEKVEAEKKAEDOKEDRRNPTNTYKTLLEIAESDVEVKAELEL 346
QY 121 VKEAEKEPRNEEKYKQAKAEVESKKAATRLKIKTRKKAEEBAKRAAEEDVKKEKPA 180
DB 347 VKEAEKEPRNEEKYKQAKAEVESKKAATRLKIKTRKKAEEBAKRAAEEDVKKEKPA 406
RESULT 11
AAU76890
ID AAU76890 standard; Protein; 693 AA.
XX
AC AAU76890;
XX
DT 05-JUN-2002 (first entry)
XX
DE Pneumococcal surface protein C19 TIGR (PspC19 TIGR).
XX
KW Hic factor H; FH; antibacterial; PspC19 TIGR;
KW pneumococcal surface protein C19 TIGR.
XX
OS Streptococcus pneumoniae.
XX
PN WO200208426-A2.
XX
PD 31-JAN-2002.
XX
PF 19-JUL-2001; 2001WO-EP08409.
XX
PR 20-JUL-2000; 2000SE-0002738.
XX
PA (HANS-) HANSA MEDICAL AB.
XX
PI Bjorck L, Sjolholm A, Janulczyk R, Pozzi G, Iannelli F;
XX WPI; 2002-257337/30.
XX

PT Polypeptide having ability to bind factor H, and proteins and peptides
PT derived from polypeptides useful as vaccine for infections caused by
PT bacteria and for identifying agents inhibiting binding of factor H to
PT protein

PS Disclosure: Page 59-61, 63pp; English.

XX The invention relates to the Streptococcus pneumoniae Hic polypeptide
CC having the ability to bind factor H (fH). The Hic protein and the
CC polynucleotide encoding it can be used to make vaccine compositions
CC capable of generating an immune response to Streptococcus pneumoniae or
CC binding to an anti-protein Hic antibody. The protein and its homologues
CC are useful for identifying an agent which inhibits binding of factor H to
CC Streptococcus pneumoniae comprising incubating any one of the
CC polypeptides with factor H and a test agent, monitoring binding of fH to
CC the polypeptide and determining thereby whether the test agent inhibits
CC binding of factor to the polypeptide. This sequence represents the
CC pneumococcal surface protein C19 TIGR (PepC19 TIGR), a factor H binding
CC protein.

XX Sequence 693 AA;

Query Match 100.0%; Score 889; DB 23; Length 693;
Best Local Similarity 100.0%; Pred. No. 2.3e-62;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IKTDREAEAEAKRRADAKGKRGVPEGLATPPDKENDAKSSDSVGEETLPS 60
DB 265 IKTDREAEAEAKRRADAKGKRGVPEGLATPPDKENDAKSSDSVGEETLPS 324
QY 61 PSLEPEKVAEAEKVEAEAKKAEDQKEEDRRNPTNTYKTLLELEIAESDVEVKAEL 120
DB 325 PSLEPEKVAEAEKVEAEAKKAEDQKEEDRRNPTNTYKTLLELEIAESDVEVKAEL 384
QY 121 VKEBAKEPRNEEKVKQAKAEVESKKAATRLLEKIKTDKKAAEEBAKRAAEEDVKKEKPA 180
DB 385 VKEBAKEPRNEEKVKQAKAEVESKKAATRLLEKIKTDKKAAEEBAKRAAEEDVKKEKPA 444

RESULT 12

ABU02720 ID ABU02720 standard; Protein; 693 AA.

XX AC ABU02720;

DT 11-FEB-2003 (first entry)

DE 5. pneumoniae type 4 strain protein from coding region #2299.

XX Bacterial meningitis; pneumonia; sepsis; otitis media;
KW ear infection; antinflammatory; antibacterial; immunostimulant;
KW auditory; respiratory; gene therapy; vaccine.

XX Streptococcus pneumoniae type 4 strain.

OS WO200277021-A2.

PN 03-OCT-2002.

PD 27-MAR-2002; 2002WO-1B02163.

PF 27-MAR-2001; 2001GB-0007656.

PR (CHIR-) CHIRON SPA.

PA (GENO-) INST GENOMIC RES.

XX Maignani V, Tettein H, Fraser C;

PI WPI, 2003-040579/03.

DR N-PSDB; ABX08011.

XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT useful as medicaments for treating or preventing a disease or infection

PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
PT or ear infection

PS Claim 1; SEQ ID No 4598; 56pp; English.

XX The invention relates to a protein comprising or having at least 50%
CC identity to any of the 2469 amino acid sequences, identified in the
CC specification (available on a computer readable format), or its fragment,
CC expressed from 2469 of 2489 identified DNA coding regions from the
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC AB556454. Also included are an antibody which binds one of the
CC proteins, treating a patient by administering the protein, DNA or
CC antibody (in a composition), a kit comprising first and second primers,
CC which are the nucleic acid cited above or fragments between nucleotides
CC 8-100 of a sequence not defined in the specification, for amplifying a
CC target sequence contained within a Streptococcus nucleic acid sequence,
CC where the first primer is substantially complementary to the target
CC sequence and the second primer is substantially complementary to the
CC complement of the target sequence, and where the parts of the primers
CC having substantial complementarity define the termini of the target
CC sequence to be amplified, assay comprising contacting a test compound
CC with the protein, and determining whether the test compound binds to the
CC protein and a Streptococcus pneumoniae bacterium, where one or more
CC genes encoding the proteins has been rendered inactive. The proteins,
CC nucleic acid molecules, antibody and compositions are useful as
CC medicaments for treating or preventing a disease or infection due to
CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,
CC sepsis, otitis media or ear infection. They are also useful in developing
CC vaccines, diagnostics and antibiotics. The methods are useful for
CC identifying immunodominant proteins. The present sequence is one of
CC the 2469 proteins expressed by the identified coding regions from the
CC genomic sequence.

CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 693 AA;

Query Match 100.0%; Score 889; DB 24; Length 693;
Best Local Similarity 100.0%; Pred. No. 2.3e-62;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IKTDREAEAEAKRRADAKGKRGVPEGLATPPDKENDAKSSDSVGEETLPS 60
DB 265 IKTDREAEAEAKRRADAKGKRGVPEGLATPPDKENDAKSSDSVGEETLPS 324
QY 61 PSLEPEKVAEAEKVEAEAKKAEDQKEEDRRNPTNTYKTLLELEIAESDVEVKAEL 120
DB 325 PSLEPEKVAEAEKVEAEAKKAEDQKEEDRRNPTNTYKTLLELEIAESDVEVKAEL 384
QY 121 VKEBAKEPRNEEKVKQAKAEVESKKAATRLLEKIKTDKKAAEEBAKRAAEEDVKKEKPA 180
DB 385 VKEBAKEPRNEEKVKQAKAEVESKKAATRLLEKIKTDKKAAEEBAKRAAEEDVKKEKPA 444

RESULT 13

AAV81653 ID AAV81653 standard; Protein; 694 AA.

XX AC AAV81653;

DT 24-MAY-2000 (first entry)

DE Streptococcus pneumoniae protein sequence ID302.

XX Streptococcus pneumoniae; vaccine; screening; protein antigen;
KW antibacterial; antinflammatory; meningitis; infection; diagnosis;
KW pneumococcal disease.

OS Streptococcus pneumoniae.

PN WO200006737-A2.

```
XX 10-FEB-2000.
PD 27-JUL-1999; 99WO-GB02451.
XX 27-JUL-1998; 98GB-0016337.
XX 27-JUL-1998; 98GB-0016337.
PR 19-MAR-1999; 99US-0125164.
XX (MCR-) MICROBIAL TECHNIQS LTD.
PA
XX Gilbert CFC, Hansbro PM,
PI WPI; 2000-195300/17.
XX New Streptococcal protein, useful as a vaccine, for diagnosis of
PT pneumococcal diseases and for screening agents capable of antagonizing
PT or inhibiting expression of the protein
XX
XX Claim 2; Page 96; 108pp; English.
PS
XX AA051501 to AA051679 represent specifically claimed protein sequences
CC isolated from Streptococcus pneumoniae. AA051407 to AA051590 represent
CC specifically claimed nucleotide sequences isolated from S. pneumoniae.
CC The sequences have antibacterial and antiinflammatory properties.
CC The protein sequences, and fragments of them, are useful as immunogens
CC and/or antigens. The nucleotide sequences can be used in vaccines and in
CC diagnostic assays. The proteins and nucleotides can be useful for the
CC detection and diagnosis of S. pneumoniae. The protein sequences are also
CC useful for screening an agent capable of antagonizing, inhibiting or
CC interfering with the function or expression of the proteins in which the
CC agent is useful for treatment or prophylaxis of S. pneumoniae infection
CC and meningitis. AA055591 to AA055614 represent primers used in the
CC exemplification of the present invention.
XX
XX Sequence 694 AA;
SQ
XX Query Match 100.0%; Score 889; DB 21; Length 694;
XX Best Local Similarity 100.0%; Pred. No. 2.3e-62;
XX Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IKTDREBAEBAEKRRADAKGQKPKGRARGVPGELATPDKKENDAKSSDSSVGEETLPS 60
DB 265 IKTDREBAEBAEKRRADAKGQKPKGRARGVPGELATPDKKENDAKSSDSSVGEETLPS 324
QY 61 PSLKPEKKVAEAEKKVEBAKKKADQKEDRRNYPNTYKTLLEIAESDVEVKKAEL 120
DB 325 PSLKPEKKVAEAEKKVEBAKKKADQKEDRRNYPNTYKTLLEIAESDVEVKKAEL 384
QY 121 VKEBAKEPRNEKVKQAQAEVSKKAEATRLKIKTRDKKAEBAKKRAAEDVKYKCPA 180
DB 385 VKEBAKEPRNEKVKQAQAEVSKKAEATRLKIKTRDKKAEBAKKRAAEDVKYKCPA 444
XX
XX RESULT 14
XX AA032189
XX ID AA032189 standard; Protein; 460 AA.
XX
XX AA032189;
XX
XX 01-FEB-2000 (first entry)
XX
XX N-terminal choline binding protein A (Cbpa) truncate.
XX
XX Choline binding protein; Cbpa; truncate; adhesin; immunogen;
XX vaccine; diagnosis; therapy; pneumococcus; sepsis; meningitis;
XX otitis media; pneumonia.
XX
XX Streptococcus pneumoniae.
XX
XX Synthetic.
XX
XX WO951188-A2.
XX
XX 14-OCT-1999.
```

```
XX 07-APR-1999; 99WO-US07669.
XX 07-APR-1998; 98US-0056019.
XX 07-APR-1998; 98US-0056019.
PR 07-APR-1998; 98US-0080878.
XX
XX (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.
XX (MEDI-) MEDIMUNE INC.
XX
XX Tuomanen EI, Maure HR, Wizemann TM, Johnson LS, Koenig S;
PI WPI; 1999-633690/54.
XX
XX New N-terminal choline binding protein A truncate polypeptides, used to
PT develop products for the diagnosis, prevention and treatment of
PT pneumococcal infections
XX
XX Claim 47; Fig 2A-B; 160pp; English.
PS
XX The present sequence represents an N-terminal choline binding
CC protein A (Cbpa) truncate, denoted Ntype 4 Cbpa trun, derived from
CC Streptococcus pneumoniae Cbpa. Claimed vaccines contain and
CC N-terminal Cbpa truncate such as the present sequence, or a
CC polypeptide comprising a conserved region of the Cbpa truncate.
CC The vaccines provide protection or elicit protective antibodies to
CC bacterial infection, specifically pneumococcus. Antibodies and
CC and antagonists against the N-terminal Cbpa truncates are used in
CC diagnosis and passive immunotherapy.
XX
XX Sequence 460 AA;
SQ
XX Query Match 99.7%; Score 886; DB 20; Length 460;
XX Best Local Similarity 99.4%; Pred. No. 2.4e-62;
XX Matches 179; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 IKTDREBAEBAEKRRADAKGQKPKGRARGVPGELATPDKKENDAKSSDSSVGEETLPS 60
DB 242 IKTDREBAEBAEKRRADAKGQKPKGRARGVPGELATPDKKENDAKSSDSSVGEETLPS 301
QY 61 PSLKPEKKVAEAEKKVEBAKKKADQKEDRRNYPNTYKTLLEIAESDVEVKKAEL 120
DB 302 PSLKPEKKVAEAEKKVEBAKKKADQKEDRRNYPNTYKTLLEIAESDVEVKKAEL 361
QY 121 VKEBAKEPRNEKVKQAQAEVSKKAEATRLKIKTRDKKAEBAKKRAAEDVKYKCPA 180
DB 362 VKEBAKEPRNEKVKQAQAEVSKKAEATRLKIKTRDKKAEBAKKRAAEDVKYKCPA 421
XX
XX RESULT 15
XX AA049140
XX ID AA049140 standard; Protein; 446 AA.
XX
XX AA049140;
XX
XX 17-JAN-2000 (first entry)
XX
XX Amino acid sequence of choline-binding protein fragment #4.
XX
XX Truncated surface binding protein; alpha helix; choline binding protein;
XX vaccine; invasive bacterial infection; otitis media; sepsis;
XX meningitis; lobar pneumonia infection; antibody; immature immune system;
XX immunocompromised.
XX
XX Streptococcus pneumoniae.
XX
XX Key Location/Qualifiers
XX Misc-difference 171
XX Misc-difference 172 /note= "Encoded by AAT"
XX Misc-difference 173 /note= "Encoded by TAC"
XX Misc-difference 174 /note= "Encoded by TTA"
XX
XX Misc-difference 174
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:28:52 ; Search time 47.3285 Seconds
(without alignments)
2786.162 Million cell updates/sec

Title: US-09-298-523C-13
Perfect score: 2537
Sequence: 1 MFASKSRKRVHYSRKFSVG.....NRLTQQQPKTEKPAQDPSTP 511

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*
15: SP virus:*
16: SP bacteriaph:*
17: SP archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2516	99.2	693	16 Q97N74	Q97N74 streptococc
2	2507	98.8	693	2 Q9KX21	Q9KX21 streptococc
3	2473.5	97.5	696	2 Q9KX32	Q9KX32 streptococc
4	2459.5	96.9	516	2 Q9R0T2	Q9R0T2 streptococc
5	2390	94.2	695	2 Q9KX36	Q9KX36 streptococc
6	2040.5	80.4	684	2 Q9KX17	Q9KX17 streptococc
7	1788	70.5	695	2 Q9KX51	Q9KX51 streptococc
8	1782	70.2	681	2 Q9KX16	Q9KX16 streptococc
9	1780.5	70.2	660	2 Q9KX18	Q9KX18 streptococc
10	1770.5	69.9	709	2 Q9KX38	Q9KX38 streptococc
11	1770.5	69.8	487	2 Q9R0T3	Q9R0T3 streptococc
12	1754.5	69.2	752	2 Q9KX30	Q9KX30 streptococc
13	1732.5	68.3	752	2 Q9KX33	Q9KX33 streptococc
14	1730	68.2	564	2 Q69188	Q69188 streptococc
15	1730	68.2	701	2 Q9KX48	Q9KX48 streptococc
16	1730	68.2	701	16 Q8DN05	Q8DN05 streptococc

17	1726	68.0	701	2 Q9R0T5	Q9R0T5 streptococc
18	1725.5	68.0	730	2 Q9KX47	Q9KX47 streptococc
19	1718	67.7	581	2 Q33742	Q33742 streptococc
20	1715.5	67.6	660	2 Q9KX41	Q9KX41 streptococc
21	1713.5	67.5	680	2 Q9KX43	Q9KX43 streptococc
22	1713.5	67.5	730	2 Q9KX31	Q9KX31 streptococc
23	1689	66.6	699	2 Q9KX42	Q9KX42 streptococc
24	1683	66.3	655	2 Q9KX50	Q9KX50 streptococc
25	1631	64.3	657	2 Q9KX29	Q9KX29 streptococc
26	1616	63.7	684	2 Q9KX46	Q9KX46 streptococc
27	1609.5	63.4	488	2 Q9R0T6	Q9R0T6 streptococc
28	1591.5	62.7	678	2 Q9KX54	Q9KX54 streptococc
29	1589.5	62.7	678	2 Q9KX49	Q9KX49 streptococc
30	1582.5	62.4	680	2 Q9KX11	Q9KX11 streptococc
31	1578	62.2	681	2 Q9KX13	Q9KX13 streptococc
32	1573	62.0	676	2 Q9KX39	Q9KX39 streptococc
33	1573	62.0	666	2 Q9KX14	Q9KX14 streptococc
34	1569.5	61.9	663	2 Q30874	Q30874 streptococc
35	1398	55.1	739	2 Q9R0T4	Q9R0T4 streptococc
36	1345	53.0	820	2 Q9R0T1	Q9R0T1 streptococc
37	1343.5	53.0	869	2 Q9KX27	Q9KX27 streptococc
38	1330	52.4	929	2 Q9ZAY5	Q9ZAY5 streptococc
39	1330	52.4	929	2 Q9KX19	Q9KX19 streptococc
40	1225	48.3	565	2 Q9KX15	Q9KX15 streptococc
41	1198.5	47.2	769	2 Q9KX40	Q9KX40 streptococc
42	1195	47.1	667	2 Q9KX28	Q9KX28 streptococc
43	1175	46.3	523	2 Q33753	Q33753 streptococc
44	1165.5	45.9	770	2 Q9KX37	Q9KX37 streptococc
45	1161.5	45.8	769	2 Q9PD01	Q9PD01 streptococc

ALIGNMENTS

RESULT 1

ID Q97N74 PRELIMINARY, PRT; 693 AA.

AC Q97N74; 01-OCT-2001 (TrEMBLrel. 18, Created)

DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Choline binding protein A.

GN SP2190.

OS Streptococcus pneumoniae.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI_TaxID=1313;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=TIGR4;

RX MEDLINE=21357209; PubMed=11463916;

RA Tetteilin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,

RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,

RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,

RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,

RA Holtzapfel E., Khouiri H., Wolf A.M., Utterback T.R., Hansen C.L.,

RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,

RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,

RT Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;

RT "Complete genome sequence of a virulent isolate of Streptococcus

pneumoniae";

RL Science 293:498-506(2001).

DR EMBL: AEO07507; AAK76241.1; -.

DR TIGR: SP2190; -.

DR InterPro: IPR002479; CW binding.

DR InterPro: IPR005877; Gpos YSIRK.

DR Pfam: PF01473; CW binding_1; 8.

DR Pfam: PF04650; YSIRK_signal; 1.

DR TIGRPFAM: TIGR01168; YSIRK_signal; 1.

KW Complete proteome.

SQ SEQUENCE 693 AA; 77762 MW; 6F8F47C32E344A41 CRC64;

Query Match 99.2%; Score 2516; DB 16; Length 693;


```
Db 61 GEOPKLDSEDRKARKEVEEYVKIIGESYAKSTKRKHTITVALVNLNINIKNEYLINKIV 120
Qy 120 ESTSSOQOILMMESRSKYDEAVSKFEKDSSSSSDSTKPEASDTAKPKPTEPGKIV 179
Db 121 ESTSSOQOILMMESRSKYDEAVSKFEKDSSSSSDSTKPEASDTAKPKPTEPGKIV 180
Qy 180 AEAKKVVEA-KKADQKEEDRRNPTTYKTLELAESDVEYKKALELVKVKANEP 238
Db 181 AEAKKVVEA-KKADQKEEDRRNPTTYKTLELAESDVEYKKALELVKVKANEP 240
Qy 239 DEQIKQAEAEVESKQAEATRLIKITDREAEAEARRADAKQGP---KGRKGV 294
Db 241 DEQIKQAEAEVESKQAEATRLIKITDREAEAEARRADAKQGP---KGRKGV 300
Qy 295 PGEATPKKENDAKSSSSVGEETLPSPSLKPEKVAEAEKVEAKKADQKEEDRR 354
Db 301 PGEATPKKENDAKSSSSVGEETLPSPSLKPEKVAEAEKVEAKKADQKEEDRR 360
Qy 355 NYPTTYKTLELAESDVEYKKALELVKKEAKEPRNEEKVKQAKAEVESKKAETRL 414
Db 361 NYPTTYKTLELAESDVEYKKALELVKKEAKEPRNEEKVKQAKAEVESKKAETRL 420
Qy 415 KIKTRKKAEEBAKKAEDVKKEKPAEOPAPAPAKPAPAKPENPAEQPKAEKP 474
Db 421 KIKTRKKAEEBAKKAEDVKKEKPAEOPAPAPAKPAPAKPENPAEQPKAEKP 480
Qy 475 ADQAEEDYARRSEEEYRLTQOOPKTEKPAOPSTP 511
Db 481 -DQAEEDYARRSEEEYRLTQOOPKTEKPAOPSTP 516
```

RESULT 4

```
Q9RQT2 PRELIMINARY; PRT; 516 AA.
ID Q9RQT2
AC Q9RQT2
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Hypothetical 57.9 kDa protein (Fragment).
GN pSpC.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=L81905;
RX MEDLINE=20038319; PubMed=10569772;
RA Brooks-Walter A., Briles D.E., Hollingshead S.K.;
RT "The pspC gene of streptococcus pneumoniae encodes a polymorphic
protein, pSpC, which elicits cross-reactive antibodies to pspC and
RT provides immunity to pneumococcal bacteremia.";
RL Infect. Immun. 67:6533-6542 (1999).
DR EMBL; AF068649; AAF13459.1; -
DR InterPro; IPR005877; Gpos_YsIRK.
DR Pfam; PF04650; YsIRK_signal; 1.
DR TIGRFAMs; TIGR01168; YsIRK_signal; 1.
DR Hypothetical protein.
KW NON TER
FT 516
SQ SEQUENCE 516 AA; 57886 MW; 1119782688BB7E35 CRC64;
```

```
Query Match 96.9%; Score 2459.5; DB 2; Length 516;
Best Local Similarity 97.9%; Pred. No. 2.8e-99;
Matches 502; Conservative 2; Mismatches 6; Indels 3; Gaps 3;
```

Qy 1 MFAKSKERKHYISIRKFSVG-ASVVVASLVGSGVHAHATENGATQVPTSSNRANESQAEQ 59
Db 1 MFAKSKERKHYISIRKFSVG-SVSVASLVGSGVHAHATENGATQVPTSSNRANESQAEQ 59

Qy 60 GEOPKLDSEDRKARKEVEEYVKIIGESYAKSTKRKHTITVALVNLNINIKNEYLINKIV 119
Db 60 GEOPKLDSEDRKARKEVEEYVKIIGESYAKSTKRKHTITVALVNLNINIKNEYLINKIV 119

```
Qy 120 ESTSESQOILMMESRSKYDEAVSKFEKDSSSSSDSTKPEASDTAKPKPTEPGKIV 179
Db 120 ESTSESQOILMMESRSKYDEAVSKFEKDSSSSSDSTKPEASDTAKPKPTEPGKIV 179
Qy 180 AEAKKVVEA-KKADQKEEDRRNPTTYKTLELAESDVEYKKALELVKVKANEP 238
Db 180 AEAKKVVEA-KKADQKEEDRRNPTTYKTLELAESDVEYKKALELVKVKANEP 239
Qy 239 DEQIKQAEAEVESKQAEATRLIKITDREAEAEARRADAKQGP---KGRKGV 298
Db 240 DEQIKQAEAEVESKQAEATRLIKITDREAEAEARRADAKQGP---KGRKGV 299
Qy 299 ATPPKKENDAKSSSSVGEETLPSPSLKPEKVAEAEKVEAKKADQKEEDRRNPT 358
Db 300 ATPPKKENDAKSSSSVGEETLPSPSLKPEKVAEAEKVEAKKADQKEEDRRNPT 359
Qy 359 NYPTTYKTLELAESDVEYKKALELVKKEAKEPRNEEKVKQAKAEVESKKAETRL 418
Db 360 NYPTTYKTLELAESDVEYKKALELVKKEAKEPRNEEKVKQAKAEVESKKAETRL 419
Qy 419 DRKAEBAKKAEDVKKEKPAEOPAPAPAKPAPAKPENPAEQPKAEKPADQ 478
Db 420 DRKAEBAKKAEDVKKEKPAEOPAPAPAKPAPAKPENPAEQPKAEKPADQ 479
Qy 479 AEDYARRSEEEYRLTQOOPKTEKPAOPSTP 511
Db 480 AEDYARRSEEEYRLTQOOPKTEKPAOPSTP 512
```

RESULT 5

```
Q9KK36 PRELIMINARY; PRT; 695 AA.
ID Q9KK36
AC Q9KK36;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Surface protein PspC.
GN pSpC.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=g394;
RA Iannelli F., Oggioni M.R., Pozzi G.;
RT "Allelic Variation in the Highly Polymorphic Locus pspC of
RT Streptococcus pneumoniae.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF154023; AAF73792.1; -
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR005877; Gpos_YsIRK.
DR Pfam; PF01473; CW_binding_1; 8.
DR Pfam; PF04650; YsIRK_signal; 1.
DR TIGRFAMs; TIGR01168; YsIRK_signal; 1.
DR SEQUENCE 695 AA; 78361 MW; F8C94F08C4EFPFB4 CRC64;
```

```
Query Match 94.2%; Score 2390; DB 2; Length 695;
Best Local Similarity 95.2%; Pred. No. 3.7e-96;
Matches 492; Conservative 6; Mismatches 11; Indels 8; Gaps 4;
```

Qy 1 MFAKSKERKHYISIRKFSVG-ASVVVASLVGSGVHAHATENGATQVPTSSNRANESQAEQ 59
Db 1 MFAKSKERKHYISIRKFSVG-SVSVASLVGSGVHAHATENGATQVPTSSNRANESQAEQ 60

Qy 60 GEOPKLDSEDRKARKEVEEYVKIIGESYAKSTKRKHTITVALVNLNINIKNEYLINKIV 119
Db 61 GEOPKLDSEDRKARKEVEEYVKIIGESYAKSTKRKHTITVALVNLNINIKNEYLINKIV 120

Qy 120 ESTSESQOILMMESRSKYDEAVSKFEKDSSSSSDSTKPEASDTAKPKPTEPGKIV 179
Db 121 ESTSESQOILMMESRSKYDEAVSKFEKDSSSSSDSTKPEASDTAKPKPTEPGKIV 180

```
QY 180 AEAKKKVEEA--KKAKDQKEEDRRNYPTTYKTLELEIAESDVEYKKALELVKVKANPR 238
DB 181 AEAKKKVEEA--KKAKDQKEEDRRNYPTTYKTLELEIAESDVEYKKALELVKVKANPR 240
QY 239 DEOKIKQAEAEVESKQAEATRLKKIKTDRREAEAEAKRADAKEOG---KPKGRAGV 294
DB 241 DEBKIKQAEAEVESKQAEATRLKKIKTDRREAEAEAKRADAKEOG---KPKGRAGV 300
QY 295 PGEIATPDKKENDAKSSDSSVGEETLPSPSLKPEKKVAEAKKVEAKKAEADQKEEDRR 354
DB 301 LGEQATPDKKENDAKSSDSSVGEETLPSPSLKPEKKVAEAKKVEAKKAEADQKEEDRR 360
QY 355 NPYNTYTTLELEIAESDVEYKKALELVKKEAKPRNEEKVKQAKAEVESKKAETLE 414
DB 361 NPYNTYTTLELEIAESDVEYKKALELVKKEAKPRNEEKVKQAKAEVESKKAETLE 420
QY 415 KIKTRKKAEEAEAKKAEDVKKEKPAEQOPAPAPAEKPAPEKPEENPAEQPAEK 474
DB 421 KIKTRKKAEEAEAKKAEDVKKEKPAEQOPAPAPAEKPAPEKPEENPAEQPAEK 478
QY 475 ADQQAEDYARRSEEEYNRLTQQQPPKTEKPAQPS 511
DB 479 ADQQAEDYARRSEEEYNRLTQQQPPKTEKPAQPS 515

RESULT 6
Q9KK17 PRELIMINARY; PRT; 684 AA.
ID 09KK17;
AC 09KK17;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Surface protein PspC.
GN PSPC.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=stre22;
RA Iannelli F., Oggioni M.R., Pozzi G.;
RT "Allelic Variation in the Highly Polymorphic Locus pspC of
RT Streptococcus pneumoniae.";
RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF154039; AAF73811.1; -.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR005877; GPos_YsIRK.
DR Pfam; PF01473; CW_binding_1; 7.
DR Pfam; PF04650; YsIRK_signal; 1.
DR TIGRFAMs; TIGR01168; YsIRK_signal; 1.
SQ SEQUENCE 684 AA; 77589 MW; 20FB892F82FDDF25 CRC64;

Query Match 80.4%; Score 2040.5; DB 2; Length 684;
Best Local Similarity 80.6%; Pred. No. 4.7e-81;
Matches 425; Conservative 34; Mismatches 49; Indels 19; Gaps 6;
```

```
QY 238 RDEOKIKQAEAEVESKQAEATRLKKIKTDRREAEAEAK-----RRADAKEGPKGRAX 291
DB 238 RDEBKIKQAEAEVESKQAEATRLKKIKTDRREAEAEAK-----RRADAKEGPKGRAX 297
QY 292 RGVGEIATPDKKENDAKSSDSSVGEETLPSPSLKPEKKVAEAKKVEAKKAEADQKEE 351
DB 298 RGVGEIATPDKKENDAKSSDSSVGEETLPSPSLKPEKKVAEAKKVEAKKAEADQKEE 357
QY 352 DRRNYPTNYTTLELEIAESDVEYKKALELVKKEAKPRNEEKVKQAKAEVESKKAET 411
DB 358 DRRNYPTNYTTLELEIAESDVEYKKALELVKKEAKPRNEEKVKQAKAEVESKKAET 417
QY 412 RLEKIKTRKKAEEAEAKKAEDVKKEKPAEQOPAPAPAEK-----PAPAPKEN 464
DB 418 RLEKIKTRKKAEEAEAKKAEDVKKEKPAEQOPAPAPAEK-----PAPAPKEN 477
QY 465 PAEQPAEKPAEQQAEDYARRSEEEYNRLTQQQPPKTEKPAQPS 511
DB 478 PAEQPAEKPAEQQAEDYARRSEEEYNRLTQQQPPKTEKPAQPS 524

RESULT 7
Q9KK51 PRELIMINARY; PRT; 695 AA.
ID 09KK51;
AC 09KK51;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Surface protein PspC.
GN PSPC.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC6307;
RA Iannelli F., Oggioni M.R., Pozzi G.;
RT "Allelic Variation in the Highly Polymorphic Locus pspC of
RT Streptococcus pneumoniae.";
RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF154009; AAF73776.1; -.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR005877; GPos_YsIRK.
DR Pfam; PF01473; CW_binding_1; 10.
DR Pfam; PF04650; YsIRK_signal; 1.
DR TIGRFAMs; TIGR01168; YsIRK_signal; 1.
SQ SEQUENCE 695 AA; 78521 MW; F208BC7285DA4FCE CRC64;

Query Match 70.5%; Score 1788; DB 2; Length 695;
Best Local Similarity 73.6%; Pred. No. 3.9e-70;
Matches 380; Conservative 39; Mismatches 51; Indels 46; Gaps 7;
```


QY 236 GELATPDKKENDAKSSDSSVGEETLPSPLKPEKVAEAKKEPRNEEKVQAKAEVSK 355
DB 260 GEPATPDKKENDAKSSDSSVGEETLPSPLKPKGKVAEAKKEPRNEEKVQAKAEVSK 319
QY 336 YPTNTYKTLLEIAESDVEVKAELELVKEAKPEPRNEEKVQAKAEVSKAETRLK 415
DB 320 YPTNTYKTLLEIAESDVEVKAELELVKEAKPEPRNEEKVQAKAEVSKAETRLK 379
QY 416 IKTDRKKAEEBAKRAAEEDKVEKPAPOPAKPAKPAKPAKPAKPAKPAKPAKPA 475
DB 380 IKTDRKKAEEBAKRAAEEDKVEKPAPOPAKPAKPAKPAKPAKPAKPAKPAKPA 439
QY 476 DQAEEDYARSRSEENRNLTOOQPKTEKPAQSTP 511
DB 440 DQAEEDYARSRSEENRNLTOOQPKTEKPAQSTP 475

RESULT 8
Q9KK16 PRELIMINARY; PRT; 681 AA.
ID Q9KK16
AC Q9KK16
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Surface protein PspC.
GN PspC.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=stf25;
RA Iannelli F., Oggioni M.R., Pozzi G.;
RT "Allelic Variation in the Highly Polymorphic Locus pspC of
RT Streptococcus pneumoniae."
RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF154040; AAF73812.1;
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR005877; Gpos_Ysirk.
DR Pfam; PF04650; Ysirk_signal; 1.
DR TIGRFAMs; TIGR01168; Ysirk_signal; 1.
SQ SEQUENCE 681 AA; 76728 MW; 38FE4782653D51A8 CRC64;

Query Match 70.2%; Score 1782; DB 2; Length 681;
Best Local Similarity 70.6%; Pred. No. 6.9e-70;
Matches 379; Conservative 43; Mismatches 51; Indels 64; Gaps 9;

QY 1 MFAKSEKRYHYSTRKPSVG-ASVVVASLVNGSVVHATENGATOVPTSSNRANESQAEQ 59
DB 1 MFAKSEKRYHYSTRKPSVG-ASVVVASLVNGSVVHATENGATOVPTSSNRANESQAEQ 57
QY 60 GEQPKLDSERDARKEVEEYVKKIVGESYAKSTKKRHTITVALVNLNINIKNELINKIV 119
DB 58 -----TEYMAAKAQVDEYITKTL-----QLDRRHIONVGLITKLGVITKTEYHLRLS 104
QY 120 ESTSESQOLIMMESRSKVDVAVSKPEKSSSSSSSSSTKPEASDTAKPNKPTPEGEKV 179
DB 105 VSKESSEAE-LPSEVKAALDAFAEQFKDT-----LPTEPKKV 142
QY 180 AEAKKYVEEA-KKAKDQKEEDRRNYPTTYKTLLEIAESDVEVKAELELVKAKAEPR 238
DB 143 AEAKKYVEEA-KKAKDQKEEDRRNYPTTYKTLLEIAESDVEVKAELELVKAKAEPR 202
QY 239 DEQKITQAEAVESKQAEATRLKKITDREAEAEAKRRADAK-----EQGPKR 286
DB 203 DEQKITQAEAVESKQAEATRLKKITDREAEAEAKRRADAK-----EQGPKR 262
QY 287 KGRKRGVPELATPDKKENDAKSSDSSVGEETLPSPLKPEKVAEAKKEPRNEEKVQAKAE 346
DB 263 KGRKRGVPELATPDKKENDAKSSDSSVGEETLPSPLKPEKVAEAKKEPRNEEKVQAKAE 322

QY 347 DQKEEDRRNYPTNTYKTLLEIAESDVEVKAELELVKEAKPEPRNEEKVQAKAEVSK 406
DB 323 DQKEEDRRNYPTNTYKTLLEIAESDVEVKAELELVKEAKPEPRNEEKVQAKAEVSK 382
QY 407 KAEATRLLEIKITDRKKAEEBAKRAAEEDKVEKPAPOPAKPAKPAKPAKPAKPA 459
DB 383 KAEATRLLEIKITDRKKAEEBAKRAAEEDKVEKPAPOPAKPAKPAKPAKPAKPA 442
QY 460 KPEPNPAQPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPA 511
DB 443 KPEPNPAQPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPA 499

RESULT 9
Q9KK18 PRELIMINARY; PRT; 660 AA.
ID Q9KK18
AC Q9KK18
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Surface protein PspC.
GN PspC.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=stf2;
RA Iannelli F., Oggioni M.R., Pozzi G.;
RT "Allelic Variation in the Highly Polymorphic Locus pspC of
RT Streptococcus pneumoniae."
RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF154038; AAF73810.1;
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR005877; Gpos_Ysirk.
DR Pfam; PF04650; Ysirk_signal; 1.
DR TIGRFAMs; TIGR01168; Ysirk_signal; 1.
SQ SEQUENCE 660 AA; 74515 MW; 374EB582FD09A659 CRC64;

Query Match 70.2%; Score 1780.5; DB 2; Length 660;
Best Local Similarity 73.4%; Pred. No. 7.8e-70;
Matches 383; Conservative 38; Mismatches 48; Indels 53; Gaps 9;

QY 1 MFAKSEKRYHYSTRKPSVG-ASVVVASLVNGSVVHATENGATOVPTSSNRANESQAEQ 59
DB 1 MFAKSEKRYHYSTRKPSVG-ASVVVASLVNGSVVHATENGATOVPTSSNRANESQAEQ 57
QY 60 GEQPKLDSERDARKEVEEYVKKIVGESYAKSTKKRHTITVALVNLNINIKNELINKIV 119
DB 58 -----TEYMAAKAQVDEYITKTL-----QLDRRHIONVGLITKLGVITKTEYHLRLS 104
QY 120 ESTSESQOLIMMESRSKVDVAVSKPEKSSSSSSSSSTKPEASDTAKPNKPTPEGEKV 179
DB 105 VSKESSEAE-LPSEVKAALDAFAEQFKDT-----LPTEPKKV 142
QY 180 AEAKKYVEEA-KKAKDQKEEDRRNYPTTYKTLLEIAESDVEVKAELELVKAKAEPR 238
DB 143 AEAKKYVEEA-KKAKDQKEEDRRNYPTTYKTLLEIAESDVEVKAELELVKAKAEPR 202
QY 239 DEQKITQAEAVESKQAEATRLKKITDREAEAEAKRRADAK-----EQGPKR 289
DB 203 DEQKITQAEAVESKQAEATRLKKITDREAEAEAKRRADAK-----EQGPKR 261
QY 290 AKRGVGEATLPDKKENDAKSSDSSVGEETLPSPLKPEKVAEAKKEPRNEEKVQAKAEVSK 349
DB 262 AKRGVGEATLPDKKENDAKSSDSSVGEETLPSPLKPEKVAEAKKEPRNEEKVQAKAEVSK 321
QY 350 EEDRRNYPTNTYKTLLEIAESDVEVKAELELVKEAKPEPRNEEKVQAKAEVSK 409
DB 322 EEDRRNYPTNTYKTLLEIAESDVEVKAELELVKEAKPEPRNEEKVQAKAEVSK 381

QY 410 ATRLEKITDPRKKAEEBAKKAEEEDKYKEKPAEOPAPAPAKKAPAPAKPENAPAPOP 469
DB 362 ATRLEKITDPRKKAEEBAKKAEEEDKYKEKPAEOPAPAPAKA--APAPKPPAPAPOP 438
QY 470 KAEKPADOAEEDYARSEEEYNRLTQOOPPKTEKPAOPSTP 511
DB 439 KAEKPADOAEEDYARSEEEYNRLTQOOPPKTEKPAOPSTP 480

RESULT 10

Q9K38 PRELIMINARY; PRT; 709 AA.
ID Q9K38
AC Q9K38;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Surface protein PspC.
GN PSpC.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RC SEQUENCE FROM N.A.
RA Iannelli F., Ovgioni M.R., Pozzi G.;
RT "Allelic Variation in the Highly Polymorphic Locus pspC of
RT Streptococcus pneumoniae."
RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF154021; AAF7390.1; -
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR005877; Gpos_YsIRK.
DR InterPro; IPR00533; Trogomyosin.
DR Pfam; PF01473; CW_binding_1; 9.
DR Pfam; PF04650; YsIRK_signal; 1.
DR PRINTS; PRO0194; TROGOMYSIN.
DR TIGRfams; TIGR01168; YsIRK_signal; 1.
SQ SEQUENCE 709 AA; 80251 MW; 32BBC96E380EBB7A CRC64;

Query Match 69.9%; Score 1773; DB 2; Length 709;
Best Local Similarity 70.8%; Pred. No. 1.8e-69;
Matches 380; Conservative 49; Mismatches 54; Indels 54; Gaps 11;
QY 1 MFASSEKRYHYSTRKESVG-ASVVVASLWGSVVAHT-ENEGATQVTSNRANESQ-- 56
DB 1 MFASNEERKHYSTRKESVGASVAVASLWGSVVAHTAEGVYSSSCSSSEAKKSQTE 60
QY 57 -----AEOGEPKLDSEBDKARKEVEEYVKIVGESYAKSTKKRTITVALVNLNINIKN 112
DB 61 HMKAAEQCKHREIDLNNKAKAIDEYIEKTLSE--IQDGRKKTQNFALNKLRIKT 118
QY 113 EYLN--KIVESTSESQLOI---LMMESRSKYDEAVSKEKSSSSSSSSSTKPEASDTA 167
DB 119 EYLVGLSVLKESESELPKAKELPEKIKELTAAEFHKDT----- 160
QY 168 KPNKPTGEGEKVAEKKKVEEA-KKAKQOEEDRRNYPTITYKTLELTAESDVEYKAE 226
DB 161 -----LRPEGEKVAEKKKVEEAKEKKAQOEEDYENYPTITYKTLELTAESDVEYKAE 215
QY 227 LELVVKANEPDRDEKIKQAEAEVESKQAEATRLKKITDREBAEBAKRRADAK----- 281
DB 216 LELVVEANKPPNEBKQKQAEKVESKQAEATRLKKITDREBAEBAKRRADAKLKA 275
QY 282 -----EQGPKRAGKGVPEGLATPDKKENDAKSSDSVGEETLPSPSLKPEKVAEA 334
DB 276 EKNATATSGQPKRAGKGVPEGLATPDKKENDAKSSDSVGEETLPSPSLKPEKVAEA 335
QY 335 EKKVBEAKKADQOEEDRRNYPTITYKTLELTAESVVEYKAELELVKEBAKPRNEE 394
DB 336 EKKVBEAKKADQOEEDRRNYPTITYKTLELTAESVVEYKAELELVKEBAKPRDEE 395
QY 395 KVKQAKAEVESKKAATLEKIKTDPRKKAEEBAKKAEEEDKYKEKPAEOPAPAPAKAE 454

DB 396 KIKQAEKVESKKAATLEKIKTDPRKKA--EBAKKAEEEDKYKEKPAEOPAPAPQPE 454
QY 455 KPAPKPPENPAEOPKAEKPADOAEEDYARSEEEYNRLTQOOPPKTEKPAOPSTP 511
DB 455 K--PAPKPEKPAEHPKAKENRADQOAEEDYARSEEEYNRLTQOOPPKTEKPAOPSTP 509

RESULT 11

Q9ROT3 PRELIMINARY; PRT; 487 AA.
ID Q9ROT3
AC Q9ROT3;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Hypothetical 55.0 kDa protein (Fragment).
GN PSpC.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RC SEQUENCE FROM N.A.
RA Brooks-Walter A., Briles D.E., Hollinghead S.K.;
RT "The pspC gene of streptococcus pneumoniae encodes a polymorphic
RT protein, PspC, which elicits cross-reactive antibodies to PspA and
RT provides immunity to pneumococcal bacteremia."
RL Infect. Immun. 67:6533-6542(1999).
DR EMBL; AF068648; AAF13458.1; -
DR InterPro; IPR005877; Gpos_YsIRK.
DR Pfam; PF04650; YsIRK_signal; 1.
DR TIGRfams; TIGR01168; YsIRK_signal; 1.
DR Hypothetical protein.
FT NON TER 487
SQ SEQUENCE 487 AA; 54962 MW; 7BDBB21C65341E86 CRC64;

Query Match 69.8%; Score 1770.5; DB 2; Length 487;
Best Local Similarity 71.9%; Pred. No. 1.6e-69;
Matches 378; Conservative 43; Mismatches 50; Indels 55; Gaps 8;
QY 1 MFASSEKRYHYSTRKESVG-ASVVVASLWGSVVAHTENEGATQVTSNRANESQAEQ 59
DB 1 MFASNEERKHYSTRKESVGASVAVASLWGSVVAHTTEKAVTTQVATSPKAKKSQ--- 57
QY 60 GEOPKLDSEBDKARKEVEEYVKIVGESYAKSTKKRTITVALVNLNINIKNEYLNKIV 119
DB 58 -----TEHMKAAQOVDEYITKTL-----QLDRRKHQVGLTGLGVITKTEYLRLS 104
QY 120 ESTSESQLOILMMSRSKVDVAESKFEKSSSSSSSSSTKPEASDTAKPKPTPEGEKV 179
DB 105 VSKESSEAE-LPSIRKAKLDAAEQFKDT-----LPTPEGKV 142
QY 180 AEAKKYVEA-KKAKQOEEDRRNYPTITYKTLELTAESDVEYKAELELVKANKANEP 238
DB 143 AEAKKYVEAKKADQOEEDRRNYPTITYKTLELTAESDVEYKAELELVKEBAKGR 202
QY 239 DEQIKQAEAEVESKQAEATRLKKIKTDREAE-----EBAKRRADAEQG 284
DB 203 NEQVNNQAKKAVESKQAEATRLKKIKTDREAEATRLLENIKTDREKAEBAKRRADAEQD 262
QY 285 KPKRAGKGVPEGLATPDKKENDAKSSDSVGEETLPSPSLKPEKVAEAEKKYVEAKK 344
DB 263 ESKRVRGVGPEQATPDKKENDAKSSDSVGEETLPSPSLKPEKVAEAEKKYVEAKK 322
QY 345 AEQDKEEDRRNYPTNTYKTLELTAESDVEYKAELELVKEBAKPRNEEKVQAKAEVE 404
DB 323 AKQDKEEDRRNYPTNTYKTLELTAESDVEYKAELELVKEBAKPRNEEKVQAKAEVE 382
QY 405 SKKAEATRLKKIKTDPRKKAEEBAKKAEEEDKYKEKPAEOPAPAPAKKAPAPAPEN 464
DB 383 SKKAEATRLKKIKTDPRKKA--EBAKKAEEEDKYKEKPAEOPAPAPQPEKAPAPAPEN 441

Qy 465 PAEPKAEKPADQAEEDYARSEEEYNRLTQOOPKTEKPAQST 510
 Db 442 PAEPKAEKPADQAEEDYARSEEEYNRLTQOOPKTEKPAQST 487

RESULT 12

Q9K30 PRELIMINARY; PRT; 752 AA.
 AC Q9K30
 DT 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE Surface protein PspC.
 GN PspC.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_Taxid=1313;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=g408;
 RA Iannelli F., Oggioni M.R., Pozzi G.;
 RT "Allelic Variation in the Highly Polymorphic Locus pspC of
 RT Streptococcus pneumoniae."
 RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF154029; AAF73798.1;
 DR InterPro: IPR002479; CM binding.
 DR InterPro: IPR005877; Gpos_YsIRK.
 DR Pfam: PF01473; CM binding_1; 12.
 DR Pfam: PF04650; YsIRK signal; 1.
 DR TIGRFAMs: TIGR01168; YsIRK signal; 1.
 SQ SEQUENCE 752 AA; 85296 MW; D3E50B8CFC2ABD3 CRC64;

Query Match 69.2%; Score 1754.5; DB 2; Length 752;
 Best Local Similarity 70.8%; Pred. No. 1.2e-68;
 Matches 379; Conservative 41; Mismatches 48; Indels 67; Gaps 11;

Qy 1 MFASKSRKHYHSIRKFSVG-ASVVVASLWGSVVAHTEGATQVPTSSNRANESQAEQ 59
 Db 1 MFASKSRKHYHSIRKFSVGASVVVASLWGSVVAHTEGATQVPTSSNRANESQAEQ 45
 Qy 60 GEOPKLDSEKDKARK--EVEEYVKKIVGESYAKTKKRTHTVALVNLNKKYEL- 115
 Db 46 ---PKVTPSGQNSKKAPDEVKSHLEKILSE--IQDKRKHQNLANKLSRIQTEYFY 100
 Qy 116 ---NKIVESTSE---SQQLIMESRSKVDEAVSKFEKSSSSSSSSSTKPEASDT 166
 Db 101 LKKKLKALSLTSKTEBELTSKTEBELTSKTEBELDAAFQFKDTLS----- 145
 Qy 167 AKPNKPTPEQBEKVAEAKKVEEA-KKAKDQKEEDRRNYPTITTYKTLELEIAESDVEVKA 225
 Db 146 ---TEPEKTKVAEAKQKVAEAKKADQKEEDRRNYPTITTYKTLELEIAEFVAKYEA 139
 Qy 226 ELLELVKKAEPREKIKIQAEEAVESKQAEATRLKKIKTDREAEAEAKRRADAKE--- 282
 Db 200 ELKLLKKAADQTRDEEIKIQAEEAVESKQAEATRLKKIKTDREAEAEAKRRADAKE 259
 Qy 283 ---OGKPRGAKRGVPGELATPDKKENDAKSSDSVGEETLPSLSLPEKKVAEAK 336
 Db 260 EEAAPSDGKPRGAKRGVPGELATPDKKENDAKSSDSVGEETLPSLSLPEKKVAEAK 319
 Qy 337 KVEBAKKAEDQKEEDRRNYPTITTYKTLELEIAESDVEVKAELVKEAEKPRNEEKV 396
 Db 320 KVEBAKKAEDQKEEDRRNYPTITTYKTLELEIAESDVEVKAELVKEAEKPRNEEKV 379
 Qy 397 KQAEAVESKKAETRLKKIKTDREAEAEAKRRADAKE---GKPRGAKRGVPG 456
 Db 380 KQAEAVESKKAETRLKKIKTDREAEAEAKRRADAKE---GKPRGAKRGVPG 438
 Qy 457 APAKPEKPAEPKAEKPADQAEEDYARSEEEYNRLTQOOPKTEKPAQSTP 511
 Db 439 -PAKPEKPAEPKAEKPADQAEEDYARSEEEYNRLTQOOPKTEKPAQSTP 492

RESULT 13

Q9K33 PRELIMINARY; PRT; 752 AA.
 AC Q9K33
 DT 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE Surface protein PspC.
 GN PspC.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_Taxid=1313;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=g40;
 RA Iannelli F., Oggioni M.R., Pozzi G.;
 RT "Allelic Variation in the Highly Polymorphic Locus pspC of
 RT Streptococcus pneumoniae."
 RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF154026; AAF73795.1;
 DR InterPro: IPR002479; CM binding.
 DR InterPro: IPR005877; Gpos_YsIRK.
 DR Pfam: PF01473; CM binding_1; 13.
 DR Pfam: PF04650; YsIRK signal; 1.
 DR TIGRFAMs: TIGR01168; YsIRK signal; 1.
 SQ SEQUENCE 752 AA; 85321 MW; E24BF1DBF50A7F8 CRC64;

Query Match 68.3%; Score 1732.5; DB 2; Length 752;
 Best Local Similarity 71.1%; Pred. No. 1e-67;
 Matches 366; Conservative 55; Mismatches 47; Indels 47; Gaps 8;

Qy 1 MFASKSRKHYHSIRKFSVG-ASVVVASLWGSVVAHTEGATQVPTSSNRANESQAEQ 59
 Db 1 MFASKSRKHYHSIRKFSVGASVVVASLWGSVVAHTEGATQVPTSSNRANESQAEQ 57
 Qy 60 GEOPKLDSEKDKARK--EVEEYVKKIVGESYAKTKKRTHTVALVNLNKKYELN 119
 Db 58 ---TEHMKAAQGVDEYVNMKI---QDKRKHQNLANKLSRIQTEYFELN 104
 Qy 120 ESTSESQQLIMESRSKVDEAVSKFEKSSSSSSSSSTKPEASDTAKNKPTEPEKV 179
 Db 105 VLEKSKKEBELTSKTEBELDAAFQFKDT-----LKPEEKV 141
 Qy 180 AEAKKVEEA-KKAKDQKEEDRRNYPTITTYKTLELEIAESDVEVKAELVVKYKAEPR 238
 Db 142 EEOQKVEEAEKKAQKEEDRRNYPTITTYKTLELEIAESDVEVKAELVVKYKAGSR 201
 Qy 239 DEQIKIQAEEAVESKQAEATRLKKIKTDREAEAEAKRRADAKEQ--GKPRGAKRGVPG 296
 Db 202 NEKKIKKAAVEESKKAETRLKKIKTDREAEAEAKRRADAKEQ--GKPRGAKRGVPG 261
 Qy 297 ELATPDKKENDAKSSDSVGEETLPSLSLPEKKVAEAKKVAEAKKADQKEEDRRNY 356
 Db 262 EEAATPDKKENDAKSSDSVGEETLPSLSLPEKKVAEAKKVAEAKKADQKEEDRRNY 321
 Qy 357 PTITTYKTLELEIAESDVEVKAELVKEAEKPRNEEKVQKAEAVESKKAETRLKKIK 416
 Db 322 PTITTYKTLELEIAESDVEVKAELVKEAEKPRNEEKVQKAEAVESKKAETRLKKIK 381
 Qy 417 KTDREAEAEAKRRADAKE---GKPRGAKRGVPG 476
 Db 382 KTDREAEAEAKRRADAKE---GKPRGAKRGVPG 437
 Qy 477 QQAEEDYARSEEEYNRLTQOOPKTEKPAQSTP 511
 Db 438 QQAEEDYARSEEEYNRLTQOOPKTEKPAQSTP 472

RESULT 14

069188

ID 069188 PRELIMINARY; PRT; 564 AA.
AC 069188;
DT 01-AUG-1998 (TREMblrel. 07, Created)
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE C3-binding protein.
GN PBCA.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OK NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Cp1200;
RX MEDLINE=20281293; PubMed=10820017;
RA Cheng Q., Finkel D., Hostetter M.K.;
RT "Novel purification scheme and functions for a C3-binding protein from
RT Streptococcus pneumoniae."
RL Biochemistry 39:5450-5457(2000).
DR EMBL: AF067128; AAC17445.1; -;
DR InterPro; IPR002479; Cw binding.
DR InterPro; IPR005877; Gpos_YsIRK.
DR Pfam; PF01473; Cw binding_1; 4.
DR Pfam; PF04650; YsIRK signal; 1.
DR TIGRFAMs; TIGR01168; YsIRK signal; 1.
SQ SEQUENCE 564 AA; 63596 MW; 884C97BAA04AED20 CRC64;

Query Match 68.2%; Score 1730; DB 2; Length 564;
Best Local Similarity 70.7%; Pred. No. 1e-67;
Matches 371; Conservative 49; Mismatches 47; Indels 58; Gaps 10;

QY 1 MPASKEKRVHYSIKFSVSG-ASVVVASLVGSGVVAHATENEGATOVPTSSNRANESQAEQ 59
DB 1 MPASKEKRVHYSIKFSVSG-ASVVVASLVGSGVVAHATENEGATOVPTSSNRANESQAEQ 54
QY 60 GEOPKKLDSEDRKARKE-VEEYVKKIIGESYAKSTKKRHTTTVALVNEINNIKNIEYLNK- 117
DB 55 -----TEHRKAAKQVVDYEYIKMLRE--IQDRKHTQVNLNITLSAIKTKYLREL 104
QY 118 -IVESTSSQOILMMESRSKVDKAVSKFEKDSSSSSSSDSTKPEASDTAKPNKPTPEG 176
DB 105 NVLEEKSKDELPL--SEIKAKLDAFAFEKFKDT-----LKPFG 138
QY 177 EKVAEAKKKVVEA-KKAKDQKEEDRRNYPTIYKTLLEIAESDVYKKAELVYKVKAN 235
DB 139 EKVAEAKKKVVEAKKAEADQKEEDRRNYPTIYKTLLEIAESDVYKKAELVYKVKAN 198
QY 236 EPRDEOKIKQAEAVESKQAEATRLKKIKTDREAEAEAKRRADAK-----EOKP 286
DB 199 ESRNEGITIKQAEVSKKAEATRLLENIKTDREKAEAEAKRRADAKLEANVATSDQCKP 258
QY 287 KGRARGVPGELATPDKKENDAKSSDSSVGEETLPSPLKPEKKVAEAEKKVVEAKKAE 346
DB 259 KGRARGVPGELATPDKKENDAKSSDSSVGEETLPSPLKPEKKVAEAEKKVVEAKKAE 318
QY 347 DQKEEDRRNYPTIYKTLLEIAESDVYKKAELVYKVKAEAEAEAEAEAEAEAEAEAE 406
DB 319 DQKEEDRRNYPTIYKTLLEIAESDVYKKAELVYKVKAEAEAEAEAEAEAEAEAEAE 378
QY 407 KAEATRLLEIKITDRKKAEEAEAKKAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 466
DB 379 KAEATRLLEIKITDRKKAEEAEAKKAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 436
QY 467 EOPKAEKPADQAEEDYARRSEEEYNRLTQOOPKTEKPAQSPSTP 511
DB 437 EOPKAEKPADQAEEDYARRSEEEYNRLTQOOPKTEKPAQSPSTP 481

RESULT 15
Q9KK48 PRELIMINARY; PRT; 701 AA.
AC Q9KK48;
DT 01-OCT-2000 (TREMblrel. 15, Created)

DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Surface protein PspC.
GN PspC.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OK NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D39;
RA Iannelli F., Oggioni M.R., Pozzi G.;
RT "Allelic Variation in the Highly Polymorphic Locus pspC of
RT Streptococcus pneumoniae."
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF154012; AAF73779.1; -;
DR InterPro; IPR002479; Cw binding.
DR InterPro; IPR005877; Gpos_YsIRK.
DR Pfam; PF01473; Cw binding_1; 10.
DR Pfam; PF04650; YsIRK signal; 1.
DR TIGRFAMs; TIGR01168; YsIRK signal; 1.
SQ SEQUENCE 701 AA; 79098 MW; A80E31FE2846F1BF CRC64;

Query Match 68.2%; Score 1730; DB 2; Length 701;
Best Local Similarity 70.7%; Pred. No. 1.3e-67;
Matches 371; Conservative 49; Mismatches 47; Indels 58; Gaps 10;

QY 1 MPASKEKRVHYSIKFSVSG-ASVVVASLVGSGVVAHATENEGATOVPTSSNRANESQAEQ 59
DB 1 MPASKEKRVHYSIKFSVSG-ASVVVASLVGSGVVAHATENEGATOVPTSSNRANESQAEQ 54
QY 60 GEOPKKLDSEDRKARKE-VEEYVKKIIGESYAKSTKKRHTTTVALVNEINNIKNIEYLNK- 117
DB 55 -----TEHRKAAKQVVDYEYIKMLRE--IQDRKHTQVNLNITLSAIKTKYLREL 104
QY 118 -IVESTSSQOILMMESRSKVDKAVSKFEKDSSSSSSSDSTKPEASDTAKPNKPTPEG 176
DB 105 NVLEEKSKDELPL--SEIKAKLDAFAFEKFKDT-----LKPFG 138
QY 177 EKVAEAKKKVVEA-KKAKDQKEEDRRNYPTIYKTLLEIAESDVYKKAELVYKVKAN 235
DB 139 EKVAEAKKKVVEAKKAEADQKEEDRRNYPTIYKTLLEIAESDVYKKAELVYKVKAN 198
QY 236 EPRDEOKIKQAEAVESKQAEATRLKKIKTDREAEAEAKRRADAK-----EOKP 286
DB 199 ESRNEGITIKQAEVSKKAEATRLLENIKTDREKAEAEAKRRADAKLEANVATSDQCKP 258
QY 287 KGRARGVPGELATPDKKENDAKSSDSSVGEETLPSPLKPEKKVAEAEKKVVEAKKAE 346
DB 259 KGRARGVPGELATPDKKENDAKSSDSSVGEETLPSPLKPEKKVAEAEKKVVEAKKAE 318
QY 347 DQKEEDRRNYPTIYKTLLEIAESDVYKKAELVYKVKAEAEAEAEAEAEAEAEAEAE 406
DB 319 DQKEEDRRNYPTIYKTLLEIAESDVYKKAELVYKVKAEAEAEAEAEAEAEAEAEAE 378
QY 407 KAEATRLLEIKITDRKKAEEAEAKKAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 466
DB 379 KAEATRLLEIKITDRKKAEEAEAKKAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 436
QY 467 EOPKAEKPADQAEEDYARRSEEEYNRLTQOOPKTEKPAQSPSTP 511
DB 437 EOPKAEKPADQAEEDYARRSEEEYNRLTQOOPKTEKPAQSPSTP 481

Search completed: November 21, 2003, 13:33:49
Job time : 51.3285 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:29:32 ; Search time 7.29378 Seconds
(without alignment)
2373.306 Million cell updates/sec

Title: US-09-298-523C-13_COPY_263_442
Perfect score: 889
Sequence: 1 IKTDREAEAEAKRRADAKE.....AEEAKRKAEEDEKVEKPA 180

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: PIR.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	889	100.0	693	2 H95255	choiline binding pr
2	803.5	90.4	701	2 H98120	choiline binding pr
3	194	21.8	1110	2 I51116	NF-180 - sea lamp
4	188.5	21.2	1020	1 QFHHU	neurofilament trip
5	186	20.9	335	2 T33457	hypothetical prote
6	183	20.6	1701	2 T09127	probable erythrocy
7	182.5	20.5	771	1 A33430	h-caldesmon - chic
8	181.5	20.4	312	2 T25994	hypothetical prote
9	181.5	20.4	1403	2 T11583	probable translati
10	180	20.2	606	2 A43427	neurofilament trip
11	178.5	20.1	332	2 B43427	neurofilament prot
12	178.5	20.1	1038	2 JCS487	clausetrin - chicke
13	177	19.9	1087	1 QFMSH	neurofilament trip
14	176.5	19.9	644	2 S55395	neurofilament prot
15	176.5	19.9	1072	1 A37221	neurofilament trip
16	176	19.8	762	2 G88436	protein T04A8.13 l
17	176	19.8	791	2 T24435	hypothetical prote
18	176	19.8	1178	2 S78475	mannosylphosphoryl
19	175	19.7	1052	1 A44937	kineoplast-associ
20	173	19.5	1390	2 S51364	sperm tail-specific
21	172.5	19.4	854	2 S02003	neurofilament trip
22	172	19.3	1132	2 T43483	translactin initia
23	171.5	19.3	805	2 E70474	translactin initia
24	170.5	19.2	344	2 S34153	met101-1 protein -
25	170.5	19.2	532	2 C96608	hypothetical prote
26	170.5	19.2	758	2 I50479	neurofilament medi
27	169.5	19.1	385	2 T19201	hypothetical prote
28	169.5	19.1	639	2 E84565	hypothetical prote
29	169	19.0	1359	2 T34036	hypothetical prote

30	168.5	19.0	407	1 EDBEQ3	immediate-early pr
31	168.5	19.0	1192	2 A71623	probable secreted
32	167	18.8	5327	2 T13564	microtubule-associ
33	166.5	18.7	845	2 A45659	neurofilament trip
34	166.5	18.7	6642	2 T29757	protein UNC-89 - C
35	164	18.4	1526	2 A45605	mature-parasite-in
36	163.5	18.3	2364	2 A56577	microtubule-associ
37	162.5	18.3	784	2 PNO009	neurofilament trip
38	162	18.2	491	2 C97267	hypothetical prote
39	160	18.0	667	2 A40713	cylicin I - bovine
40	159.5	17.9	210	2 T28771	hypothetical prote
41	159.5	17.9	721	2 A33319	microtubule-associ
42	159	17.9	849	2 S00030	neurofilament trip
43	158.5	17.8	619	2 A41971	surface protein ps
44	158.5	17.8	619	2 A97887	surface protein ps
45	158.5	17.8	806	2 T23648	hypothetical prote

ALIGNMENTS

```
RESULT 1
H95255
choiline binding protein A [imported] - Streptococcus pneumoniae (strain TIGR4)
C/Species: Streptococcus pneumoniae
C/Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C/Accession: H95255
R/Retcelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzaple,
nson, T.; Hickey, E.K.; Holt, I.E.
A/Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A./Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A/Reference number: A95000; MUID:21357209; PMID:11463916
A/Accession: H95255
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-693 <KUR>
A/Cross-references: GB:AE005672; PIDN:AAK76241.1; PID:g14973701; GSPDB:GN00164; TIGR:SPA
A/Experimental source: strain TIGR4
C/Genetics:
A/Gene: SP2190

Query Match      100.0%; Score 889; DB 2; Length 693;
Best Local Similarity 100.0%; Pred. No. 2.6e-38;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 IKTDREAEAEAKRRADAKEOGKPKGRAGVGEIATPPKENDAKSSDSVGEETLPS 60
        |||||||
Db      265 IKTDREAEAEAKRRADAKEOGKPKGRAGVGEIATPPKENDAKSSDSVGEETLPS 324

QY      61 PSLKPEKVAEAEKVEAKKAEDKEDRRNYPNTYKTLLEIAESDVEYKAELEL 120
        |||||||
Db      325 PSLKPEKVAEAEKVEAKKAEDKEDRRNYPNTYKTLLEIAESDVEYKAELEL 384

QY      121 VKEAKEPRNEEKVAKAEVESKKAEDATRLKIKTKDKKAEAEAKKAEDKVEKPA 180
        |||||||
Db      385 VKEAKEPRNEEKVAKAEVESKKAEDATRLKIKTKDKKAEAEAKKAEDKVEKPA 444

RESULT 2
H98120
choiline binding protein A [imported] - Streptococcus pneumoniae (strain R6)
C/Species: Streptococcus pneumoniae
C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C/Accession: H98120
R/Hoskins, J.A.; Albom Jr., W.; Arnold, J.; Blaszcak, L.; Bunge, S.; Dehoff, B.S.; E
e, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.W.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A/Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A./Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A/Reference number: A97872; MUID:21429245; PMID:11544234
```

A/Accession: H98120
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-701 <KBP>
A/Cross-references: GB:AB007317; PIDN:AA00797.1; PID:g15459699; GSPDB:GN00174
C/Genetics:
A/Gene: pspc

Query Match 90.4%; Score 803.5; DB 2; Length 701;
Best Local Similarity 86.8%; Pred. No. 5,4e-34;
Matches 164; Conservative 12; Mismatches 4; Indels 9; Gaps 1;

Qy 1 IKTRDEAEAEAKRRADAK-----EOKPKGRAGVPGELATPDKKENDAKSSDS 51
Db IKTRDKKAEAEAKRRADAKLKEANVATSDQKPKRAGVPGELATPDKKENDAKSSDS 285
Qy 52 SVGEETLPSPSLKPEKKYAEAEAKYEEAKKAEODKEDRRNYPNTYKTLLEIAESDV 111
Db SVGEETLPSSSLKSGKKYAEAEAKYEEAKKAEODKEDRRNYPNTYKTLLEIAESDV 345
Qy 112 EVKKAELVKEAEKPERNEEKVQAKAEVSKKAEATRLKIKTRDKKAEAEAKRAE 171
Db 346 VKKAELELVKEAEKPERDEEKIKQAKAKVSKKAEATRLKIKTRDKKAEAEAKRAE 405
Qy 172 EDKVKEXPA 180
Db 406 EDKVKEXPA 414

RESULT 3
151116
NF-180 - sea lamprey
C/Species: Petromyzon marinus (sea lamprey)
C/Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C/Accession: 151116
R/Jacobs, A.J.; Kamholz, J.; Seizer, M.E.
Brain Res. Mol. Brain Res. 29, 43-52, 1995
A/Title: The single lamprey neurofilament subunit (NF-180) lacks multiphosphorylation re
A/Reference number: 151116; MUID:95287814; PMID:7770000
A/Accession: 151116
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-1110 <JAC>
A/Cross-references: EMBL:U19361; NID:G632548; PIDN:AAA60106.1; PID:G632549
C/Superfamily: neurofilament triplet H protein

Query Match 21.8%; Score 194; DB 2; Length 1110;
Best Local Similarity 29.6%; Pred. No. 0.0052;
Matches 58; Conservative 38; Mismatches 74; Indels 26; Gaps 5;

Qy 2 KTDREAEAEAKRRADAKQKPKRAGVPGELATPDKKENDAKSSDSVGEETLPSP 61
Db 527 EKEKEAEAEAEVEAEAEETAEAAEAEAEAEAE-----EAEAEAEAEAEAEAEVEE 580
Qy 62 SLKEKKYAEAEAKYEEAKKAEODKEDRRNYPNTYKTLLEIAESDVVKKAELV 121
Db 581 IEXKAEAEAKYEEAE 639
Qy 122 KEAEKPR-NEEKVQAKAEVSKKAEATRL-----EKIKTRDKKAEAEAKRA 168
Db 640 EEEAEAEAEVTSKKATQAEVEEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAEAE 699
Qy 169 A-----AEEDVKEX 178
Db 700 ADAEAEAEAEAEAEVEE 715

RESULT 4
QFHU
neurofilament triplet H protein - human
N/Alternate names: neurofilament protein, 112k
C/Species: Homo sapiens (man)
C/Date: 30-Jun-1991 #sequence_revision 30-Jun-1991 #text_change 16-Jun-2000

C/Accession: S00979
R/Lees, J.F.; Shneidman, P.S.; Skuntz, S.F.; Carden, M.J.; Lazzarini, R.A.
EMBO J. 7, 1947-1955, 1988
A/Title: The structure and organization of the human heavy neurofilament subunit (NF-H)
A/Reference number: S00979; MUID:86328981; PMID:318108
A/Accession: S00979
A/Molecule type: DNA
A/Residues: 1-1020 <LEE>
A/Cross-references: EMBL:X15306; NID:935028; PIDN:CAA33366.1; PID:g1841430
A/Note: it is uncertain whether Met-1 or Met-2 is the initiator
C/Genetics:
A/Gene: GDB:NEFH
A/Cross-references: GDB:120225; OMIM:162230
A/Map position: 22q12.1-22q13.1
A/Introns: 295/1; 361/3; 403/2
C/Superfamily: neurofilament triplet H protein
C/Keywords: coiled coil; heterotrimer; intermediate filament; nerve; phosphoprotein
F:1-100/Domain: amino-terminal <NTD>
F:101-410/Domain: rod #status predicted <ROD>
F:411-1020/Domain: carboxyl-terminal <CTD>
F:502-826/Region: 14-residue repeats
F:503,511,518,526,532,540,546,552,560,566,574,580,586,594,600,606,614,620,628,634,640,644
(covalent) #status predicted
F:732,768/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 21.2%; Score 188.5; DB 1; Length 1020;
Best Local Similarity 30.9%; Pred. No. 0.0092;
Matches 58; Conservative 28; Mismatches 79; Indels 23; Gaps 6;

Qy 5 REAEAEAKRRADAKQKPKRAGVPGELATPDKKENDAKSSDSVGEETLPSPSLK 64
Db 699 KEAEKPEKAKSPVKEAEKPEKAKSPVKEAEKPEKAKSPVKEAKS-----PEKAKS 752
Qy 65 PEKVAAAEKVEAEAKKAEODKEDRRNYPNTYKTLLEIAESD-----VEVKAE 117
Db 753 PE-KAKTLVDVSPAKTPAKAEASPADKPEKAKSPVKEAKSPVKEAKSPVKADAKAPE 811
Qy 118 LEL-VKEAEKPERNE-----KVQAKAEVSKKAEATRLKIKTRDKKAEAEAKRAA 170
Db 812 KEIKKAEKVEKSPVKEAEKPEKAEKPEKAEKPEKAEKAEKATPTEKKSOK--EAPKKEA 869
Qy 171 EEDVKEX 178
Db 870 PKPVEEK 877

RESULT 5
T33457
hypochemical protein F36H12.3 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C/Accession: T33457
R/Blanchard, M.; Bradshaw, H.; Steliyes, L.
submitted to the EMBL Data Library, July 1998
A/Description: The sequence of C. elegans cosmid F36H12.
A/Reference number: 221346
A/Accession: T33457
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-335 <BLA>
A/Cross-references: EMBL:AF078790; PIDN:AAC26930.1; GSPDB:GN00022; CESP:F36H12.3
A/Experimental source: strain Bristol N2; clone F36H12
C/Genetics:
A/Gene: CESP:F36H12.3
A/Map position: 4
A/Introns: 32/3; 227/1; 270/2

Query Match 20.9%; Score 186; DB 2; Length 335;
Best Local Similarity 31.1%; Pred. No. 0.0047;
Matches 56; Conservative 31; Mismatches 77; Indels 16; Gaps 5;
Qy 2 KTDREAEAEAKRRADAKQKPKRAGVPGELATPDKKE--NDAKSSDSVGEETLP 59
Db 118 LEL-VKEAEKPERNE-----KVQAKAEVSKKAEATRLKIKTRDKKAEAEAKRAA 170
Db 812 KEIKKAEKVEKSPVKEAEKPEKAEKPEKAEKPEKAEKAEKATPTEKKSOK--EAPKKEA 869
Qy 171 EEDVKEX 178
Db 870 PKPVEEK 877

```

Db      55 KSKKEEKEEPKKEEKEEKSK-KSKGKSDKKKEEKPKEEKEKEPKKEEKEKKEDDK 113
Qy      60 SPULPEKKVAAEKKVEAKKKAEDKEDKEDBRNRYNTNYTKTLELIAEDVYKAELE 119
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      114 KDEKDEKDEKDEKDEKDEKDEKDEKDDKEDKDDKDKS---DNKDKDDEKKEEK 170
Qy      120 LVKEAEKPRNDEKXKQAKAEVSKKAEKTRLEKITDKKAAEEEAQRKAAEEDKYKEP 179
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      171 DDKEEKEEKEEKEEKK--EKKKE-----KKEEKEEKEEKEEKEDEKYEKPPR 220

```

RESULT 6

Probable erythrocyte-binding protein MAEBL - Plasmodium yoelii
 C/Species: Plasmodium yoelii
 C/Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
 C/Accession: T09127
 R/Kapre, S.H.I.; Noe, A.R.; Fraser, T.S.; Blair, P.L.; Adams, J.H.
 Proc. Natl. Acad. Sci. U.S.A. 95, 12302-1235, 1998
 A/Title: A family of chimeric erythrocyte binding proteins of malaria parasites.
 A/Reference number: Z16577; M01D:98115903; PMID:9448314
 A/Accession: T09127
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-1701 <KAP>
 A/Cross-references: EMBL:AF031886; NID:G2947227; P1D:G2947228
 A/Experimental source: subspecies yoelii, strain YM
 C/Genetics:
 A/Genes: maeb1
 A/Introns: 62/1, 1674/1, 1674/2, 1697/1
 C/Keywords: alternative splicing; cell binding; erythrocyte invasion

Query Match	20.6%	Score 183	DB 2	Length 1701
Best Local Similarity	29.6%	Pred. No. 0.027		
Matches 69	Conservative 40	Mismatches 66	Indels 58	Gaps 10

[illegible]

RESULT 7

h-caldesmon chicken
N:Alternate names: caldesmon, smooth muscle, calmodulin- and actin-binding protein
C:Species: Gallus gallus (chicken)
C:Date: 27-Feb-1990 #sequence, revision 27-Jun-1994 #next change 22-Jun-1999
C:Accession: A33430; A32642; A32445; A41064; A60461; PC2003; EX0022
R:Hayashi, K.; Kanda, K.; Kimizuka, F.; Kato, I.; Sobue, K.
Biochem. Biophys. Res. Commun. 164, 503-511, 1989
A:Title: Primary structure and functional expression of h-caldesmon complementary DNA
A:Reference number: A33430; PMID:90026426; PMID:2803315
A:Accession: A33430
A:Molecule type: mRNA
A:Residues: 1-771 <NAV>
A:Cross-references: GB:M28417; NID:g211895; PIDN:AAA8810.1; PID:g211896
A:Experimental source: gizzard
A:Note: part of this sequence was confirmed by protein sequencing
R:Brayan, J.; Imai, M.; Dee, R.; Moore, F.; Cook, R.G.; Lin, W.G.
J. Biol. Chem. 264, 13873-13879, 1989
A:Title: Cloning and expression of a smooth muscle caldesmon.

A:Reference number: A32642; MUID:89340480; PMID:2760048
A:Accession: A32642
A:Molecule type: mRNA
A:Residues: 1-318,334-771 <BRY>
A:Cross-references: GB:J04968; NID:g212656; PIDN:AAA49067.1; PID:g212657
A>Note: The authors translated the codon GAA for residue 743 as Ieu
A>Note: this alternative splice form is a high molecular weight caldesmon (h-caldesmon)
R:Hayashi, K.; Yamada, S.; Kanda, K.; Kimizuka, F.; Kato, I.; Sobue, K.
Biochem. Biophys. Res. Commun. 161, 38-45, 1989
A>Title: 35kDa fragment of h-caldesmon conserves two consensus sequences of the tropomyo
A:Reference number: A33445; MUID:89273666; PMID:2730665
A:Accession: A33445
A:Molecule type: mRNA
A:Residues: 466-771 <HAN>
A:Cross-references: GB:M26684; NID:g211897; PIDN:AAA4811.1; PID:g211898
R:Maik, A.S.; Carpenter, M.; Smillie, L.B.; Wang, J.H.
J. Biol. Chem. 266, 19971-19975, 1991
A>Title: Phosphorylation of caldesmon by p34(cdc2) kinase. Identification of phosphoryla
A:Reference number: A41064; MUID:92041815; PMID:1939059
A:Accession: A41064
A:Molecule type: protein
A:Residues: 597-600,678-696,711-721 <MAK>
R:Wang, C.L.A.; Wang, L.W.C.; Lu, R.C.
Biochem. Biophys. Res. Commun. 162, 746-752, 1989
A>Title: Caldesmon has two calmodulin-binding domains.
A:Reference number: A60461; MUID:89334885; PMID:2757638
A:Accession: A60461
A:Molecule type: protein
A:Residues: 2-17,'X',19-38,466-485 <WAN>
R:Hatuna, M.; Hayashi, K.; Yano, H.; Takeuchi, O.; Sobue, K.
Biochem. Biophys. Res. Commun. 197, 145-153, 1993
A>Title: Common structural and expresional properties of vertebrate caldesmon_genes.
A:Reference number: PC2003; MUID:94071934; PMID:8250919
A:Accession: PC2003
A:Molecule type: DNA
A:Residues: 74-419 <HAR>
R:Takagi, T.; Yazawa, M.; Ueno, T.; Suzuki, S.; Yegi, K.
J. Biochem. 106, 778-783, 1989
A>Title: Amino acid sequence studies on cyanogen bromide peptides of chicken caldesmon w
A:Reference number: PX0022; MUID:90130380; PMID:2613684
A:Accession: PX0022
A:Molecule type: protein
A:Residues: 462-477,'D',479-563,674-762,'A',763-771 <TAK>
C:Comment: This protein plays a vital role in the regulation of smooth muscle and nonmus
C:Comment: The binding of caldesmon to F-actin is modulated by calcium and calmodulin.
C:Comment: Two calmodulin molecules can bind to nonoverlapping domains of each caldesmon
C:keywords: actin binding; alternative splicing; calmodulin binding; cytoskeleton; muscl
F:1-771/Product: h-caldesmon #status predicted <HMAT>
F:1-118,334-771/Product: h-caldesmon, alternative splice form #status predicted <LMAT>
F:266-390/Region: 13-residue repeats
F:511-580/Region: tropomyosin binding
F:622-636/Region: tropomyosin binding
F:597,662,717/Binding site: phosphate (Thr) (covalent) (by cdc2 kinase) #status experimental
F:688,711/Binding site: phosphate (Thr) (covalent) (by cdc2 kinase) #status experimental

Query Match 20.5%; Score 182.5; DB 1; Length 771;
Best Local Similarity 33.3%; Pred. No. 0.015;
Matches 61; Conservative 28; Mismatches 71; Indels 23; Gaps 6;

OY 5 REAEAEKKRADAEQCKPKGRAGVIGELATP-----KKENDASSDSVGEEET 57
DB 171 RODEEERBEKEEKSEEEKPK-----EVPTENQDVAVAEKSTDKCEVETKTLLAVAN 225
OY 58 LPSPSLKREKVAA-AEKKVEEAKKKAEDOKEDRRNYNPNTYKTLLELAESDVEVKKA 116
DB 226 DTNMLBEQSITDAADDEKEBAKERKLTAEBEKRLAAEEKKAAEEQKAEKKAA 285
OY 117 ELELVKEBAK---EBRNKEKYQAQAEVESKKAATRLKETIKTRKKAEEBK----RKA 169
DB 286 E--EREBAKAEERKRAAEERERAKAEERKAAEERERAKAEERERAKAEERKAAEERKA 342
OY 170 AEE 172

Db 343 AEE 345

RESULT 8

T25994

hypothetical protein ZK354.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T25994

R:Johnson, D.; Wamsley, P.; Bradshaw, H.

Submitted to the EMBL Data Library, February 1997

A:Description: The sequence of C. elegans cosmid ZK354.

A:Reference number: Z20120

A:Accession: T25994

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-312 <JOH>

A:Cross-references: EMBL:U88172; PIDN:ABA42258.1; GSPDB:GN00022; CESP:ZK354.3

A:Experimental source: strain Bristol N2; clone ZK354

C:Genetics:

A:Gene: CESP:ZK354.3

A:Map position: 4

A:introns: 41/1; 226/3

Query Match 20.4%; Score 181.5; DB 2; Length 312;
Best Local Similarity 29.8%; Pred. No. 0.0075;
Matches 62; Conservative 34; Mismatches 63; Indels 49; Gaps 9;

Qy 2 KTDREAEAEARRADAEQKPKGRARGVPGELATPDKKENDAKSSDSVGEETLPS 61

Db 119 KSERDKKDF-RKEDDKKEENKESKDEBKDDV--KDKKE-DEKMDKKRGEKEEKE 174

Qy 62 SLKPE-KKVAEAKKVEBAK-----KAE-----DQKEDRR 92

Db 175 EFKEMKKEEKEKEKPEPKNDAPKKEGETGEVKKDALVNPITVEMSDRDEKKEBK 234

Qy 93 NYPNTTYTLLEIAESDVEVKKALELVKEAKPRRDEKVKQAK--AEVSKKAEATR 150

Db 235 DDKD-----EKDEKKEAKKEKEKEKEKEKEDDKEDDKKSKATK 284

Qy 151 LEKTKDRKAE-----EAKRKAEEDK 174

Db 285 SEDKSDKDKTEKSKDEKKNKSEBK 312

RESULT 9

T11583

probable translation initiation factor - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999

C:Accession: T11583

R:Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.

submitted to the EMBL Data Library, April 1996

A:Reference number: Z17295

A:Accession: T11583

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1403 <BAR>

A:Cross-references: EMBL:Z73099; NID:g1314152; PID:e241757

A:Experimental source: strain 972h(-)

C:Genetics:

A:Map position: IR

A:Note: SPAC17C9.03

Query Match 20.4%; Score 181.5; DB 2; Length 1403;
Best Local Similarity 31.3%; Pred. No. 0.027;
Matches 56; Conservative 35; Mismatches 57; Indels 31; Gaps 7;

Qy 5 REAEAEARRADAEQKPKGRARGVPGELATPDKKENDAKSSDSVGEETLPS 64

Db 562 KRBAEQAARAEAEKAEAEKAEAE--EAKAEAEENAKREAEAEKAEAE----- 613

Qy 65 PEKVAEAEKVE-EAKKAEADQEDRRNPTNTYKTLLEIAESDVEVKKAELEVKE 123

Db 614 EAKAREAEAEKAEAEKAEAEKAEAEKAEAEKAEAEKAEAEKAEAEKAEAEKAE 659

Qy 124 EAKEPRNEKVKQAKAEV-ESKKAATRLKIKTD-----RKAEAEARRAEEDVK 176

Db 660 EAENAKREAEAEKAEAEAEAEKAEAE--EKVKETEENAKRKAEAEQKREADKNPEIK 716

RESULT 10

A43427

neurofilament triplet H1 protein - rabbit (fragment)

C:Species: Oryctolagus cuniculus (domestic rabbit)

C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 10-Dec-1999

C:Accession: A43427

R:Soppet, D.R.; Beasley, L.L.; Willard, M.B.

J. Biol. Chem. 267, 17354-17361, 1992

A:Title: Evidence for unequal crossing over in the evolution of the neurofilament polype

A:Reference number: A43427; MUID:92381055; PMID:1512270

A:Accession: A43427

A:Molecule type: DNA

A:Residues: 1-606 <SOP>

A:Cross-references: GB:M94315; NID:g164990; PIDN:AAA57152.1; PID:g601930

A:Experimental source: brain

A:Note: sequence extracted from NCBI backbone (NCBIN:112010, NCBIP:112011)

C:Superfamily: neurofilament triplet H protein

C:Keywords: coiled coil; heterotrimer; intermediate filament; nerve; phosphoprotein

Query Match 20.2%; Score 180; DB 2; Length 606;
Best Local Similarity 31.9%; Pred. No. 0.016;
Matches 67; Conservative 23; Mismatches 70; Indels 50; Gaps 9;

Qy 5 REAEAEARRADAEQKPKGRARGVPGELATPDKKENDAKSSDSVGEETLPS 64

Db 307 KEAEAEAEKAEAEKAEAEKAEAEKAEAEKAEAEKAEAEKAEAEKAEAEKAE 363

Qy 65 PEKVAEAEKVEAEKKA--EDQEDRRNPTNTYKTLLEIAESDVEVKKAELEVKE 123

Db 364 EAKSPKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAE 412

Qy 124 EAKSP-----RNEKVKQ-----AKAEVSKKAEATRLKIK----- 155

Db 413 EAKSPKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAE 472

Qy 156 TDRKKAEEAEKAKA-----EEDVKYKEK 178

Db 473 ABAKDEAEADKKAAPAKMEGKEAEKPEK 502

RESULT 11

B43427

neurofilament protein H form H2 (repetitive region) - rabbit (fragment)

C:Species: Oryctolagus cuniculus (domestic rabbit)

C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 10-Jul-1998

C:Accession: B43427

R:Soppet, D.R.; Beasley, L.L.; Willard, M.B.

J. Biol. Chem. 267, 17354-17361, 1992

A:Title: Evidence for unequal crossing over in the evolution of the neurofilament polype

A:Reference number: A43427; MUID:92381055; PMID:1512270

A:Accession: B43427

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-932 <SOP>

A:Experimental source: brain

A:Note: sequence extracted from NCBI backbone (NCBIP:112012)

C:Superfamily: neurofilament triplet H protein

Query Match 20.1%; Score 178.5; DB 2; Length 332;
Best Local Similarity 30.6%; Pred. No. 0.011;
Matches 59; Conservative 30; Mismatches 79; Indels 25; Gaps 7;

Qy 6 REAEAEARRADAEQKPKGRARGV-----PGELATPDKKENDAKSSDSVGEETLPS 61

QY 1 IKTDEEAEEAKRRADAKGCKPKRAKRGVPGELATPD--KKENDAKSSDSSVGEEITLP 59
Db 268 VSVKEEKEEBAEAGEEEOEAEEVAAMAKS-PVATTPETKEEKEEKEEEOEEEBE 326
QY 60 SPSLK-----PEKKVAEAKKEVEEA---KKKAEDQKEE--DRNNY 94
Db 327 DEGVASDQAEBSGSEKSSKNKEGEOEGEETAEAGEVEEAKEKEKTEKSEEEVAAKEE 386
QY 95 PTNTYKTELETAESDVEYKKAELVLVEAKEPPNEEKVKQAKAEVBSKKA--EATLE 152
Db 387 PTEAKVGPEKAKSPVPSPVVEYKPKAEATAGGEOKEEKEEKEEKKAKESPRKE 446
QY 153 KIK-----TDRKKAEEAKRAAEEDKVKEXP 179
Db 447 KVEKEEKPKDVPKKKASPVKEEAEEATITKP 481

RESULT 15

A37221
neurofilament triplet H protein - rat
C/Species: Rattus norvegicus (Norway rat)
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Dec-1999
C/Accession: A37221; A25649; A30796; A32757; B25649
R/Chin, S.S.M.; Liem, R.K.H.
J. Neurosci. 10, 3714-3726, 1990
A/Title: Transfected rat high-molecular-weight neurofilament (NF-H) coassembles with vit
A/Reference number: A37221; MUID:91038277; PMID:2230956
A/Accession: A37221
A/Status: preliminary; not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 1-1072 <CHI>
A/Cross-references: GB:AF031879; NID:G2642597; PIDN:AAB87068.1; PID:G2642598
R/Robinson, P.A.; Wion, D.; Anderton, B.H.
FEBS Lett. 209, 203-205, 1986
A/Title: Isolation of a cDNA for the rat heavy neurofilament polypeptide (NF-H).
A/Reference number: A25649; MUID:87080760; PMID:2878828
A/Accession: A25649
A/Molecule type: mRNA
A/Residues: 230-318;472-542 <ROB>
A/Cross-references: GB:M37227
R/Dautigny, A.; Pham-Dinh, D.; Rouseel, C.; Felix, J.M.; Nussbaum, J.L.; Jolles, P.
Biochem. Biophys. Res. Commun. 154, 1099-1106, 1988
A/Title: The large neurofilament subunit (NF-H) of the rat: cDNA cloning and in situ det
A/Reference number: A30796; MUID:88309090; PMID:2457365
A/Accession: A30796
A/Molecule type: mRNA
A/Residues: 266-421, 'T', 423-427, 'T', 429-542, 'V', 556-566, 'E', 568-613, 'A', 615-725, 'S', 727-
A/Cross-references: GB:M21964; NID:G205685; PIDN:AAA41695.1; PID:G205686
R/Lieberburg, I.; Spinner, N.; Snyder, S.; Anderson, J.; Goldhaber, D.; Smulowitz, M.; C
Proc. Natl. Acad. Sci. U.S.A. 86, 2463-2467, 1989
A/Title: Cloning of a cDNA encoding the rat high molecular weight neurofilament peptide
22.
A/Reference number: A32757; MUID:89184647; PMID:2928342
A/Accession: A32757
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 559-566, 'E', 568-967, 'V', 969-997, 'GST', 1001-1022, 'E', 1024-1072 <LIB>
A/Cross-references: GB:J04517; NID:G205679; PIDN:AAA41692.1; PID:G205680
C/Superfamily: neurofilament triplet H protein
C/Keywords: coiled coil; heterotrimer; intermediate filament; nerve; phosphoprotein

Query Match 19.9%; Score 176.5; DB 1; Length 1072;
Best Local Similarity 28.0%; Pred. No. 0.039; Indels 53; Gaps 6;
Matches 59; Conservative 28; Mismatches 71;

QY 11 EAKRRADAKGCKPKRAKRGVPGELATPDKKENDAKSSDSSVGEEITLPSPLKPEKKVA 70
Db 735 EAKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEKSPAEK 794
QY 71 EAKKVEAKKKAEDQKEEDRRNYPNTYKTELETAESDVEYKKAELVLVEAKEPPNEEK 128
Db 795 KPPAEVKSPEKAKSPMKKEAK---SPEAKTLDVKSPEAKTP-----AKKEAKRPAD 843

QY 129 -RNEEKVKQ-AKAEVSKKAETRLKTI-----KTDRKAAEE----- 164
Db 844 IRSPQVKSPEKAKSPKEKSPKEETTEKVAPEKKEVKSPEVEYKAKEPEKPEKTEPATP 903
QY 165 -----AKKAAEDKVKEXP 179
Db 904 KTEVSKKDEAPKKAQPKAEKEKPLEKP 934

Search completed: November 21, 2003, 13:34:30
Job time : 8.29378 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:28:52 ; Search time 16.6715 Seconds
(without alignments)
2786.162 Million cell updates/sec

Title: US-09-298-523C-13_COPY_263_442
Perfect score: 889
Sequence: 1 IKTDREAEAEAKRADAKE.....ABEAKRAEDKVEKPA 180

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriaph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	889	100.0	516	2 Q9ROT2	Q9ROT2 streptococc
2	889	100.0	693	16 Q97A74	Q97A74 streptococc
3	880	99.0	693	2 Q9KK12	Q9KK12 streptococc
4	871	98.0	696	2 Q9KK32	Q9KK32 streptococc
5	866.5	97.5	752	2 Q9KK30	Q9KK30 streptococc
6	816.5	91.8	655	2 Q9KK30	Q9KK30 streptococc
7	815.5	91.7	680	2 Q9KK43	Q9KK43 streptococc
8	814.5	91.6	680	2 Q9KK11	Q9KK11 streptococc
9	809	91.0	660	2 Q9KK18	Q9KK18 streptococc
10	807	90.8	730	2 Q9KK47	Q9KK47 streptococc
11	804	90.4	695	2 Q9KK36	Q9KK36 streptococc
12	803.5	90.4	564	2 Q69188	Q69188 streptococc
13	803.5	90.4	581	2 Q33742	Q33742 streptococc
14	803.5	90.4	663	2 Q30874	Q30874 streptococc
15	803.5	90.4	701	2 Q9KK48	Q9KK48 streptococc
16	803.5	90.4	701	16 Q8DN05	Q8DN05 streptococc

17	803	90.3	681	2 Q9KK16	Q9KK16 streptococc
18	799.5	89.9	701	2 Q9ROT5	Q9ROT5 streptococc
19	795	89.4	730	2 Q9KK31	Q9KK31 streptococc
20	794.5	89.4	657	2 Q9KK29	Q9KK29 streptococc
21	793.5	89.3	709	2 Q9KK38	Q9KK38 streptococc
22	787	88.5	681	2 Q9KK13	Q9KK13 streptococc
23	779	87.6	678	2 Q9KK49	Q9KK49 streptococc
24	776	87.3	660	2 Q9KK41	Q9KK41 streptococc
25	772.5	86.9	695	2 Q9KK51	Q9KK51 streptococc
26	770.5	86.7	659	2 Q9KK42	Q9KK42 streptococc
27	761	85.6	684	2 Q9KK17	Q9KK17 streptococc
28	757.5	85.2	684	2 Q9KK46	Q9KK46 streptococc
29	752	84.6	487	2 Q9ROT3	Q9ROT3 streptococc
30	751.5	84.5	752	2 Q9KK33	Q9KK33 streptococc
31	728.5	81.9	488	2 Q9ROT6	Q9ROT6 streptococc
32	716.5	80.6	676	2 Q9KK39	Q9KK39 streptococc
33	716.5	80.6	696	2 Q9KK14	Q9KK14 streptococc
34	709.5	79.8	739	2 Q9ROT4	Q9ROT4 streptococc
35	704.5	79.2	869	2 Q9KK27	Q9KK27 streptococc
36	702.5	79.0	678	2 Q9KK54	Q9KK54 streptococc
37	693.5	78.0	565	2 Q9KK15	Q9KK15 streptococc
38	684.5	77.0	820	2 Q9ROT1	Q9ROT1 streptococc
39	684.5	77.0	929	2 Q9ZAV5	Q9ZAV5 streptococc
40	684.5	77.0	929	2 Q9KK19	Q9KK19 streptococc
41	551.5	62.0	523	2 Q33753	Q33753 streptococc
42	521	58.6	107	2 Q93MU1	Q93MU1 streptococc
43	517	58.2	107	2 Q93MU4	Q93MU4 streptococc
44	516.5	58.1	667	2 Q9KK28	Q9KK28 streptococc
45	514	57.8	539	2 Q33741	Q33741 streptococc

ALIGNMENTS

RESULT 1

Q9ROT2 PRELIMINARY; PRT; 516 AA.
ID Q9ROT2
AC Q9ROT2
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical 57.9 kDa protein (Fragment).
GN pspC.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=181905;
RX MEDLINE=20038319; PubMed=10569772;
RA Brooks-Walter A., Biles D.E., Hollingshead S.K.;
RT "The pspC gene of streptococcus pneumoniae encodes a polymorphic
RT protein, PspC, which elicits cross-reactive antibodies to PspA and
RT provides immunity to pneumococcal bacteremia.";
RL Infect. Immun. 67:6533-6542(1999).
DR EMBL: AF068649; AAF13459.1;
DR InterPro: IPR005877; Gpos_YSTRK.
DR Pfam: PFO4650; YSTRK_signal; 1.
DR TIGRFAMs: TIGR01168; YSTRK_signal; 1.
KW Hypothetical protein.
FT NON TER 516
SQ SEQUENCE 516 AA; 57886 MW; 1119782688B7E35 CRC64;

DEC

Query Match 100.0%; Score 889; DB 2; Length 516;
Best Local Similarity 100.0%; Pred. No. 2.4e-42;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IKTDREAEAEAKRADAKEQKPKRAKGVGELATPPKKNDAKSSNSVGEETLPS 60
DB 264 IKTDREAEAEAKRADAKEQKPKRAKGVGELATPPKKNDAKSSNSVGEETLPS 323
QY 61 PSLKPKKVAEAKKVEAKKAEDQKEDRRNYPNTYTLLEIAESVVEVKAELEL 120

Db 324 PSLEKKEVAAEAEKVEAKKAEADQKEDRRNRYTNTYKLELEIASDVEVKAELEL 383
Qy 121 VKEAKEPRNEBKVKQAKAEVSKAEATRLKIKTDKKAEEBAKRAAEEDKVKEXPA 180
Db 384 VKEAKEPRNEBKVKQAKAEVSKAEATRLKIKTDKKAEEBAKRAAEEDKVKEXPA 443

RESULT 2

Q97N74 PRELIMINARY; PRT; 693 AA.
ID Q97N74
AC Q97N74
DT 01-OCT-2001 (Tremblrel. 18, Created)
DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Choline binding protein A.
GN SP2190.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
NX NCBI_TaxID=1313;
RN [1]

SEQUENCE FROM N.A.
RA MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heijlberg J., Deboy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwin W., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radue D.,
RA Holtzaple E., Khouri H., Wolf A.M., Utecherback T.R., Hansen C.L.,
RA McDonald L.A., Feldlyum T.V., Angiolini S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RA "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001).
DR EMBL; AE007507; AAK76241.1; -
DR TIGR; SP2190; -
DR InterPro; IPR002479; CM binding.
DR InterPro; IPR003877; Gpos_YsIRK.
DR Pfam; PF01473; CM binding_1; 8.
DR Pfam; PF04650; YsIRK signal_1.
DR TIGRFAMs; TIGR01168; YsIRK_signal; 1.
KW Complete proteome.
SQ SEQUENCE 693 AA; 77762 MW; 6F8F47C32E344A41 CRC64;

Query Match 100.0%; Score 889; DB 16; Length 693;
Best Local Similarity 100.0%; Pred. No. 3.2e-42;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 IKTDREAEAEAKRRADAKGQKPKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPS 60
Db 265 IKTDREAEAEAKRRADAKGQKPKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPS 324
Qy 61 PSLEKKEVAAEAEKVEAKKAEADQKEDRRNRYTNTYKLELEIASDVEVKAELEL 120
Db 325 PSLEKKEVAAEAEKVEAKKAEADQKEDRRNRYTNTYKLELEIASDVEVKAELEL 384
Qy 121 VKEAKEPRNEBKVKQAKAEVSKAEATRLKIKTDKKAEEBAKRAAEEDKVKEXPA 180
Db 385 VKEAKEPRNEBKVKQAKAEVSKAEATRLKIKTDKKAEEBAKRAAEEDKVKEXPA 444

RESULT 3

Q9KK21 PRELIMINARY; PRT; 693 AA.
ID Q9KK21
AC Q9KK21
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Surface protein PspC.
GN PspC.
OS Streptococcus pneumoniae.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
NX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=g9;
RA Iannelli F., Oggioni M.R., Pozzi G.;
RT "Allelic Variation in the Highly Polymorphic Locus pspC of
RT Streptococcus pneumoniae.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF154035; AAF73807.1; -
DR InterPro; IPR002479; CM binding.
DR InterPro; IPR005877; Gpos_YsIRK.
DR Pfam; PF01473; CM binding_1; 8.
DR Pfam; PF04650; YsIRK signal_1.
DR TIGRFAMs; TIGR01168; YsIRK_signal; 1.
SQ SEQUENCE 693 AA; 77789 MW; 8DA1BA115665667 CRC64;

Query Match 99.0%; Score 880; DB 2; Length 693;
Best Local Similarity 99.4%; Pred. No. 1e-41;
Matches 179; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 IKTDREAEAEAKRRADAKGQKPKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPS 60
Db 265 IKTDREAEAEAKRRADAKGQKPKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPS 324
Qy 61 PSLEKKEVAAEAEKVEAKKAEADQKEDRRNRYTNTYKLELEIASDVEVKAELEL 120
Db 325 PSLEKKEVAAEAEKVEAKKAEADQKEDRRNRYTNTYKLELEIASDVEVKAELEL 384
Qy 121 VKEAKEPRNEBKVKQAKAEVSKAEATRLKIKTDKKAEEBAKRAAEEDKVKEXPA 180
Db 385 VKEAKEPRNEBKVKQAKAEVSKAEATRLKIKTDKKAEEBAKRAAEEDKVKEXPA 444

RESULT 4

Q9KK32 PRELIMINARY; PRT; 696 AA.
ID Q9KK32
AC Q9KK32
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Surface protein PspC.
GN PspC.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
NX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=g402;
RA Iannelli F., Oggioni M.R., Pozzi G.;
RT "Allelic Variation in the Highly Polymorphic Locus pspC of
RT Streptococcus pneumoniae.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF154027; AAF73796.1; -
DR InterPro; IPR002479; CM binding.
DR InterPro; IPR005877; Gpos_YsIRK.
DR Pfam; PF01473; CM binding_1; 8.
DR Pfam; PF04650; YsIRK signal_1.
DR TIGRFAMs; TIGR01168; YsIRK_signal; 1.
SQ SEQUENCE 696 AA; 78444 MW; DFAA15A16166D7D9 CRC64;

Query Match 98.0%; Score 871; DB 2; Length 696;
Best Local Similarity 97.3%; Pred. No. 3.2e-41;
Matches 179; Conservative 0; Mismatches 1; Indels 4; Gaps 1;

Qy 1 IKTDREAEAEAKRRADAKGQKPKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPS 60
Db 265 IKTDREAEAEAKRRADAKGQKPKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPS 324
Qy 57 TLPSPLPEKKEVAAEAEKVEAKKAEADQKEDRRNRYTNTYKLELEIASDVEVKAE 116

Db 325 TLPSPLKEPKVAEAEKKEAEKKAADQKEEDRRNYPNTYKTLLEIAESDVEVKKA 384
Qy 117 ELEIYKEAEKPEKKEKVAEAEKKEAEKKAATRLLEKIKTRKKAEAEKKAEDVK 176
Db 385 ELEIYKEAEKPEKKEKVAEAEKKEAEKKAATRLLEKIKTRKKAEAEKKAEDVK 444
Qy 177 EKPA 180
Db 445 EKPA 448

RESULT 5

Q9KK30 PRELIMINARY; PRT; 752 AA.
AC Q9KK30;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Surface protein PspC.
GN PspC.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=g408;
RA Iannelli F., Oggioni M.R., Pozzi G.;
RT "Allelic Variation in the Highly Polymorphic Locus pspC of
RT Streptococcus pneumoniae."
RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF154029; AAF73798.1; -
DR InterPro: IPR002479; CW binding.
DR InterPro: IPR005877; Gpos_YsIRK.
DR Pfam: PF01473; CW_binding_1; 12.
DR Pfam: PF04650; YsIRK_signal; 1.
DR TIGRFAMs: TIGR01168; YsIRK_signal; 1.
SQ SEQUENCE 752 AA; 85296 MW; D3B50B8CFC2ABD3 CRC64;

Query Match 97.5%; Score 866.5; DB 2; Length 752;
Best Local Similarity 94.2%; Pred. No. 6.1e-41;
Matches 178; Conservative 1; Mismatches 1; Indels 9; Gaps 1;

Qy 1 IKTDREAEAEKRRADAK-----QGKPKGAKRGVPELATPPDKKENDAKSSDS 51
Db 237 IKTDREAEAEKRRADAKKEAEADBDQKPKGRAGVPELATPPDKKENDAKSSDS 296
Qy 52 SVGEETLPSPSLKEPKVAEAEKKEAEKKAEDQKEEDRRNYPNTYKTLLEIAESDV 111
Db 297 SVGEETLPSPSLKEPKVAEAEKKEAEKKAEDQKEEDRRNYPNTYKTLLEIAESDV 356
Qy 112 EVKAEIYKEAEKPEKKEKVAEAEKKEAEKKAATRLLEKIKTRKKAEAEKKAEDVK 171
Db 357 EVKAEIYKEAEKPEKKEKVAEAEKKEAEKKAATRLLEKIKTRKKAEAEKKAEDVK 416
Qy 172 EDKYEKPA 180
Db 417 EDKYEKPA 425

RESULT 6

Q9KK50 PRELIMINARY; PRT; 655 AA.
AC Q9KK50;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Surface protein PspC.
GN PspC.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=8R1;
RA Iannelli F., Oggioni M.R., Pozzi G.;
RT "Allelic Variation in the Highly Polymorphic Locus pspC of
RT Streptococcus pneumoniae."
RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF154010; AAF73777.1; -
DR InterPro: IPR002479; CW binding.
DR InterPro: IPR005877; Gpos_YsIRK.
DR InterPro: IPR00533; Tropomyosin.
DR Pfam: PF01473; CW_binding_1; 8.
DR Pfam: PF04650; YsIRK_signal; 1.
DR PRINTS: PR00194; TROPOMYOSIN.
DR TIGRFAMs: TIGR01168; YsIRK_signal; 1.
SQ SEQUENCE 655 AA; 73951 MW; 30623F1EDB9D05C2 CRC64;

Query Match 91.8%; Score 816.5; DB 2; Length 655;
Best Local Similarity 93.3%; Pred. No. 3.3e-38;
Matches 168; Conservative 6; Mismatches 5; Indels 1; Gaps 1;

Qy 1 IKTDREAEAEKRRADAKQGKPKGAKRGVPELATPPDKKENDAKSSDSVGEETLPS 60
Db 228 IKTERKAEAEKRRADAKQGKPKGAKRGVPEQATPPDKKENDAKSSDSVGEETLPN 287
Qy 61 PSLLKEPKVAEAEKKEAEKKAEDQKEEDRRNYPNTYKTLLEIAESDVEVKAEIYL 120
Db 288 PSLLKEPKVAEAEKKEAEKKAEDQKEEDRRNYPNTYKTLLEIAESDVEVKAEIYL 347
Qy 121 VKEAEKPEKKEKVAEAEKKEAEKKAATRLLEKIKTRKKAEAEKKAEDVKKEKPA 180
Db 348 VKEAEKSRNEKKEKVAEAEKKEAEKKAATRLLEKIKTRKKAEAEKKAEDVKKEKPA 406

RESULT 7

Q9KK43 PRELIMINARY; PRT; 680 AA.
AC Q9KK43;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Surface protein PspC.
GN PspC.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=g376;
RA Iannelli F., Oggioni M.R., Pozzi G.;
RT "Allelic Variation in the Highly Polymorphic Locus pspC of
RT Streptococcus pneumoniae."
RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF154016; AAF73784.1; -
DR InterPro: IPR002479; CW binding.
DR InterPro: IPR005877; Gpos_YsIRK.
DR Pfam: PF01473; CW_binding_1; 8.
DR Pfam: PF04650; YsIRK_signal; 1.
DR TIGRFAMs: TIGR01168; YsIRK_signal; 1.
SQ SEQUENCE 680 AA; 76727 MW; AAD51C2A31495B8E CRC64;

Query Match 91.7%; Score 815.5; DB 2; Length 680;
Best Local Similarity 88.4%; Pred. No. 3.8e-38;
Matches 167; Conservative 10; Mismatches 3; Indels 9; Gaps 1;

Qy 1 IKTDREAEAEKRRADAK-----EGKPKGAKRGVPELATPPDKKENDAKSSDS 51
Db 237 IKTDREAEAEKRRADAKKEAEADBDQKPKGRAGVPELATPPDKKENDAKSSDS 296
Qy 52 SVGEETLPSPSLKEPKVAEAEKKEAEKKAEDQKEEDRRNYPNTYKTLLEIAESDV 111
Db 297 SVGEETLPSPSLKEPKVAEAEKKEAEKKAEDQKEEDRRNYPNTYKTLLEIAESDV 356

QY 112 EVKKAELVKEAKEPRNEEKVKAKAEESSKAAEATRLKIKTDKKAEEAKRKAAE 171
DB 357 KVKKAELVKEAKEPRNEEKVKAKAEESSKAAEATRLKIKTDKKAEEAKRKAAE 416
QY 172 EDKVKKEPA 180
DB 417 EDKVKKEPA 425

RESULT 8
Q9KK11 PRELIMINARY; PRT; 680 AA.
AC 09KK11: 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Surface protein PspC.
GN PspC.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RA Iannelli F., Oggioni M.R., Pozzi G.;
RT "Allelic Variation in the Highly Polymorphic Locus pspC of
RT Streptococcus pneumoniae."
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF154045; AAF73822.1; -
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR005877; Gpos_YsIRK.
DR Pfam; PF01473; CW_binding_1; 9.
DR Pfam; PF04650; YsIRK_signal; 1.
DR TIGRFAMs; TIGR01168; YsIRK_signal; 1.
SQ SEQUENCE 680 AA; 76185 MW; 07D682193DD0A0F9 CRC64;

Query Match 91.6%; Score 814.5; DB 2; Length 680;
Best Local Similarity 91.4%; Pred. No. 4,4e-38;
Matches 171; Conservative 3; Mismatches 4; Indels 9; Gaps 2;

QY 1 IKTDREAEAEKARRADAK-----QGKPKGRAGVPGELATPDKKENDAKSSDS 52
DB 222 IKTDREKAEAEKARRADALQENAVATSGDESKRAKGVPELATPDKKENDAKSSDS 281
QY 53 VGEETLPSPSLKPEKKVAEAEKKVAAEDQKEEDRRNYPNTYKTLELEIAESDV 112
DB 282 VGEETLPSPSLKPEKKVAEAEKKVAAEDQKEEDRRNYPNTYKTLELEIAESDV 341
QY 113 VKKAELVKEAKEPRNEEKVKAKAEESSKAAEATRLKIKTDKKAEEAKRKAAE 172
DB 342 VKKAELVKEAKEPRNEEKVKAKAEESSKAAEATRLKIKTDKKAEEAKRKAAE 400
QY 173 DKVKKEP 179
DB 401 DKVKKEP 407

RESULT 9
Q9KK18 PRELIMINARY; PRT; 660 AA.
AC 09KK18: 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Surface protein PspC.
GN PspC.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=srf2;
RA Iannelli F., Oggioni M.R., Pozzi G.;
RT "Allelic Variation in the Highly Polymorphic Locus pspC of
RT Streptococcus pneumoniae."
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF154038; AAF73810.1; -
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR005877; Gpos_YsIRK.
DR Pfam; PF01473; CW_binding_1; 8.
DR Pfam; PF04650; YsIRK_signal; 1.
DR TIGRFAMs; TIGR01168; YsIRK_signal; 1.
SQ SEQUENCE 660 AA; 74515 MW; 374EB582FD09A659 CRC64;

Query Match 91.0%; Score 809; DB 2; Length 660;
Best Local Similarity 91.0%; Pred. No. 8.6e-39;
Matches 172; Conservative 1; Mismatches 6; Indels 10; Gaps 2;

QY 1 IKTDREAEAEKARRADAK-----EQKPKGRAGVPGELATPDKKENDAKSSDS 51
DB 227 IKTDREKA-EEAKARRADAKLQENAVATSEQDSKRRANREVIGELATPDKKENDAKSSDS 285
QY 52 SVGEETLPSPSLKPEKKVAEAEKKVAAEDQKEEDRRNYPNTYKTLELEIAESDV 111
DB 286 SVGEETLPSPSLKPEKKVAEAEKKVAAEDQKEEDRRNYPNTYKTLELEIAESDV 345
QY 112 EVKKAELVKEAKEPRNEEKVKAKAEESSKAAEATRLKIKTDKKAEEAKRKAAE 171
DB 346 EVKKAELVKEAKEPRNEEKVKAKAEESSKAAEATRLKIKTDKKAEEAKRKAAE 405
QY 172 EDKVKKEPA 180
DB 406 EDKVKKEPA 414

RESULT 10
Q9KK47 PRELIMINARY; PRT; 730 AA.
AC 09KK47: 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Surface protein PspC.
GN PspC.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RA Iannelli F., Oggioni M.R., Pozzi G.;
RT "Allelic Variation in the Highly Polymorphic Locus pspC of
RT Streptococcus pneumoniae."
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF154013; AAF73780.1; -
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR005877; Gpos_YsIRK.
DR Pfam; PF01473; CW_binding_1; 11.
DR Pfam; PF04650; YsIRK_signal; 1.
DR TIGRFAMs; TIGR01168; YsIRK_signal; 1.
SQ SEQUENCE 730 AA; 82304 MW; D3CCSD0BB4E8D74E CRC64;

Query Match 90.8%; Score 807; DB 2; Length 730;
Best Local Similarity 87.0%; Pred. No. 1.2e-37;
Matches 167; Conservative 7; Mismatches 6; Indels 12; Gaps 1;

QY 1 IKTDREAEAEKARRADAK-----EQKPKGRAGVPGELATPDKKENDAKS 48
DB 225 IKTDREKAEAEKARRAEKLEKAVEKNAVATSEQKPKGRAGVPGEDATPDKKENDAKS 284
QY 49 SDSSVGEETLPSPSLKPEKKVAEAEKKVAAEDQKEEDRRNYPNTYKTLELEIAE 108
DB 285 SDSSVGEETLPSPSLKPEKKVAEAEKKVAAEDQKEEDRRNYPNTYKTLELEIAE 344

```
QY 109 SDVEYKAELELVKEAEKPRNEEKVQAKAEVESKKAATRLKIKTDKKAEEBAERK 168
DB 345 SDVYKAELELVKEAEKPRNEEKVQAKAEVESKKAATRLKIKTDKKAEEBAERK 404
QY 169 AAEDKVEKPA 180
DB 405 AAEDKVEKPA 416

RESULT 11
QYK36 PRELIMINARY; PRT; 695 AA.
ID 09K36
AC 09K36;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE Surface protein PspC.
GN PspC.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=g394;
RA Lannelli F., Oggioni M.R., Pozzi G.;
RT "Allelic Variation in the Highly Polymorphic Locus pspC of
RT Streptococcus pneumoniae."
RL Submitted (May-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF154023; AAF73792.1; -.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR005877; Gpos_YsIRK.
DR Pfam; PF01473; CW_binding_1; 8.
DR Pfam; PF04650; YsIRK_signal; 1.
DR TIGRFAMs; TIGR01168; YsIRK_signal; 1.
SQ SEQUENCE 695 AA; 78361 MW; F8C84F08C4EFFFB4 CRC64;

Query Match 90.4%; Score 804; DB 2; Length 695;
Best Local Similarity 91.3%; Pred. No. 1.7e-37;
Matches 168; Conservative 3; Mismatches 9; Indels 4; Gaps 1;

QY 1 IKTDREBAEEBAKRRADAK-----KPKGRKRGVPGELATPPKKENDAKSSDS 56
DB 265 IKTDREKAEGAKRRADAKEDDESKRKRYKGDGEOATPPKKENDAKSSDSVGE 324
QY 57 TLPSPSLKPEKKVAEAEKKVEAKKKAEDQKEEDRRNYPNTYTKTLELEIAESDV 116
DB 325 TLPSPSLKPEKKVAEAEKKVEAKKKAEDQKEEDRRNYPNTYTKTLELEIAESDV 384
QY 117 ELELVKEAEKPRNEEKVQAKAEVESKKAATRLKIKTDKKAEEBAERK 176
DB 385 ELELVKEAEKPRNEEKVQAKAEVESKKAATRLKIKTDKKAEEBAERK 444
QY 177 EKPA 180
DB 445 EKPA 448

RESULT 12
069188 PRELIMINARY; PRT; 564 AA.
ID 069188
AC 069188;
DT 01-AUG-1998 (TREMblrel. 07, Created)
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE C3-binding protein.
GN pCA.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=CP1200;
RX MEDLINE=20281293; PubMed=10820017;
RA Cheng Q., Finkel D., Hostetter W.K.;
RT "Novel purification scheme and functions for a C3-binding protein from
RT Streptococcus pneumoniae."
RL Biochemistry 39:5450-5457(2000).
DR EMBL; AF067128; AAC17445.1; -.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR005877; Gpos_YsIRK.
DR Pfam; PF01473; CW_binding_1; 4.
DR Pfam; PF04650; YsIRK_signal; 1.
DR TIGRFAMs; TIGR01168; YsIRK_signal; 1.
SQ SEQUENCE 564 AA; 63596 MW; 884C97BA04AED20 CRC64;

Query Match 90.4%; Score 803.5; DB 2; Length 564;
Best Local Similarity 86.8%; Pred. No. 1.5e-37;
Matches 164; Conservative 12; Mismatches 4; Indels 9; Gaps 1;

QY 1 IKTDREBAEEBAKRRADAK-----EOGKPKGRKRGVPGELATPPKKENDAKSSDS 51
DB 226 IKTDKKAEEBAKRRADAKLEAVNAVATSDQKPKGRKRGVPGELATPPKKENDAKSSDS 285
QY 52 SVGEETLPSPSLKEKKVAEAEKKVEAKKKAEDQKEEDRRNYPNTYTKTLELEIAESDV 111
DB 286 SVGEETLPSPSLKGKGVAEAEKKVEAEKKAEDQKEEDRRNYPNTYTKTLELEIAESDV 345
QY 112 EVKAELELVKEAEKPRNEEKVQAKAEVESKKAATRLKIKTDKKAEEBAERK 171
DB 346 KVEAELELVKEAEKPRNEEKVQAKAEVESKKAATRLKIKTDKKAEEBAERK 405
QY 172 EDKVEKPA 180
DB 406 EDKVEKPA 414

RESULT 13
033742 PRELIMINARY; PRT; 581 AA.
ID 033742
AC 033742;
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE SpA protein.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=type 47;
RX MEDLINE=98010350; PubMed=9350867;
RA Hammerschmidt S., Talay S.R., Brandtzaeg P., Chatawal G.S.;
RT "SpA, a novel pneumococcal surface protein with specific binding to
RT Immunoglobulin A and secretory component."
RL Mol. Microbiol. 25:1113-1124(1997).
DR EMBL; AU002055; CAA05159.1; -.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR005877; Gpos_YsIRK.
DR Pfam; PF01473; CW_binding_1; 4.
DR Pfam; PF04650; YsIRK_signal; 1.
DR TIGRFAMs; TIGR01168; YsIRK_signal; 1.
SQ SEQUENCE 581 AA; 65671 MW; 9CF32F2BE4490DFC CRC64;

Query Match 90.4%; Score 803.5; DB 2; Length 581;
Best Local Similarity 86.8%; Pred. No. 1.5e-37;
Matches 164; Conservative 12; Mismatches 4; Indels 9; Gaps 1;

QY 1 IKTDREBAEEBAKRRADAK-----EOGKPKGRKRGVPGELATPPKKENDAKSSDS 51
DB 226 IKTDKKAEEBAKRRADAKLEAVNAVATSDQKPKGRKRGVPGELATPPKKENDAKSSDS 285
QY 52 SVGEETLPSPSLKEKKVAEAEKKVEAKKKAEDQKEEDRRNYPNTYTKTLELEIAESDV 111
```

```

DB 286 SVGEETLPSSLSKSGKVAEAEKVAEAKKADQKEDRRNYPTNTYKTLDEIAESDV 345
QY 112 EVKKAELVLVEEAEKPEPNEEKVKQAKAEVSKAEATRLKIKTDKKAEAEAKRAAE 171
DB 346 KVEKAELVLVEEAEKPEPDEEKIKQAKAKVESKAEATRLLENIKTDKKAEAEAKRAAE 405
QY 172 EDKVEKPA 180
DB 406 EDKVEKPA 414

```

RESULT 14

```

O30874 PRELIMINARY; PRT; 663 AA.
AC O30874;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Choline binding protein A (Fragment).
GN CBPA.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OX Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R6x;
RA Rozenow C., Ryan P., Weiser J.N., Johnson S., Fontan P., Ortqvist A.,
RA Masure H.R.;
RT "Contribution of novel choline-binding proteins to adherence,
RT colonization and immunogenicity of Streptococcus pneumoniae.";
RL Mol. Microbiol. 0:0-0(1997).
DR EMBL; AF019904; AAB70838.1;
DR InterPro; IPR002479; CW_binding.
DR Pfam; PF01473; CW_binding_1; 10.
FT NON TER
SQ SEQUENCE 663 AA; 75064 MW; AFD25A7AB2ACE7B CRC64;

```

Query Match 90.4%; Score 803.5; DB 2; Length 663;
 Best Local Similarity 86.8%; Pred. No. 1.7e-37;
 Matches 164; Conservative 12; Mismatches 4; Indels 9; Gaps 1;

```

QY 1 IKTDEEAEAEKRRADAK-----EKGKPKGRAGVPGELATPDKKENDAKSSDS 51
DB 188 IKTDRKKAEEBAKRAKADKLKEANVATSDQKPKGRAGVPGELATPDKKENDAKSSDS 247
QY 52 SVGEETLPSSLSKSGKVAEAEKVAEAKKAEADQKEDRRNYPTNTYKTLDEIAESDV 111
DB 248 SVGEETLPSSLSKSGKVAEAEKVAEAKKADQKEDRRNYPTNTYKTLDEIAESDV 307
QY 112 EVKKAELVLVEEAEKPEPNEEKVKQAKAEVSKAEATRLKIKTDKKAEAEAKRAAE 171
DB 308 KVEKAELVLVEEAEKPEPDEEKIKQAKAKVESKAEATRLLENIKTDKKAEAEAKRAAE 367
QY 172 EDKVEKPA 180
DB 368 EDKVEKPA 376

```

RESULT 15

```

O9KK48 PRELIMINARY; PRT; 701 AA.
AC O9KK48;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Surface protein PspC.
GN PspC.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OX Streptococcus.
OX NCBI_TaxID=1313;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D39;
RA Iannelli F., Oggioni M.R., Pozzi G.;
RT "Allelic Variation in the Highly Polymorphic Locus pspC of
RT Streptococcus pneumoniae.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF154012; AAF73779.1;
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR005877; Gpos_YsIRK.
DR Pfam; PF01473; CW_binding_1; 10.
DR Pfam; PF04650; YsIRK_signal; 1.
DR TIGRfams; TIGR01168; YsIRK_signal; 1.
SQ SEQUENCE 701 AA; 79098 MW; A80E31FE2846F1BF CRC64;

```

Query Match 90.4%; Score 803.5; DB 2; Length 701;
 Best Local Similarity 86.8%; Pred. No. 1.8e-37;
 Matches 164; Conservative 12; Mismatches 4; Indels 9; Gaps 1;

```

QY 1 IKTDEEAEAEKRRADAK-----EKGKPKGRAGVPGELATPDKKENDAKSSDS 51
DB 226 IKTDRKKAEEBAKRAKADKLKEANVATSDQKPKGRAGVPGELATPDKKENDAKSSDS 285
QY 52 SVGEETLPSSLSKSGKVAEAEKVAEAKKAEADQKEDRRNYPTNTYKTLDEIAESDV 111
DB 286 SVGEETLPSSLSKSGKVAEAEKVAEAKKADQKEDRRNYPTNTYKTLDEIAESDV 345
QY 112 EVKKAELVLVEEAEKPEPNEEKVKQAKAEVSKAEATRLKIKTDKKAEAEAKRAAE 171
DB 346 KVEKAELVLVEEAEKPEPDEEKIKQAKAKVESKAEATRLLENIKTDKKAEAEAKRAAE 405
QY 172 EDKVEKPA 180
DB 406 EDKVEKPA 414

```

Search completed: November 21, 2003, 13:33:49
 Job time : 16.6715 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:30:37 ; Search time 21.4457 Seconds
(Without alignments)
1008.166 Million cell updates/sec

Title: US-09-298-523C-13

Perfect score: 2537

Sequence: 1 MFASKSERKRVHYIRKFSVG.....NRLTQQQPPKTEKPAQPSRP 511

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents AA:
1: /cgn2_6/prodata/1/aa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/aa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/aa/6A_COMB.pep:*
4: /cgn2_6/prodata/1/aa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/1/aa/backfilest1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2194.5	86.5	446	4	US-09-286-981B-9 Sequence 9, Appl
2	2177.5	85.8	446	4	US-09-286-981B-6 Sequence 6, Appl
3	1728	68.1	564	3	US-09-308-022-6 Sequence 6, Appl
4	1512.5	59.6	406	4	US-09-286-981B-18 Sequence 18, Appl
5	1506	57.4	631	3	US-08-847-065-25 Sequence 25, Appl
6	1455.5	57.1	426	4	US-09-286-981B-12 Sequence 12, Appl
7	1447.5	57.1	425	4	US-09-286-981B-13 Sequence 13, Appl
8	1446.5	57.0	424	4	US-09-286-981B-14 Sequence 14, Appl
9	1440.5	56.8	428	4	US-09-286-981B-7 Sequence 7, Appl
10	1432	56.4	425	4	US-09-286-981B-11 Sequence 11, Appl
11	1412.5	55.7	412	4	US-09-286-981B-17 Sequence 17, Appl
12	1410.5	55.6	419	4	US-09-286-981B-15 Sequence 15, Appl
13	1405.5	55.4	414	4	US-09-286-981B-16 Sequence 16, Appl
14	1401.5	55.2	414	4	US-09-286-981B-10 Sequence 10, Appl
15	1395	55.0	431	4	US-09-286-981B-3 Sequence 3, Appl
16	1373	54.1	453	3	US-08-961-083-38 Sequence 38, Appl
17	1373	54.1	453	4	US-09-536-784-38 Sequence 38, Appl
18	1291	50.9	413	4	US-09-286-981B-5 Sequence 5, Appl
19	1207	47.6	864	4	US-08-714-741-40 Sequence 40, Appl
20	1157	45.6	588	4	US-08-714-741-42 Sequence 42, Appl
21	904	35.6	605	4	US-08-714-741-46 Sequence 46, Appl
22	856.5	33.8	8991	4	US-08-714-741-32 Sequence 32, Appl
23	764	30.1	1231	4	US-08-714-741-41 Sequence 41, Appl
24	752	29.6	623	4	US-08-714-741-47 Sequence 47, Appl
25	714.5	28.2	251	4	US-09-286-981B-4 Sequence 4, Appl
26	636	25.1	142	3	US-08-847-065-21 Sequence 21, Appl
27	560	22.1	219	4	US-09-286-981B-8 Sequence 8, Appl

28	542	21.4	114	4	US-09-286-981B-19 Sequence 19, Appl
29	525.5	20.7	110	3	US-08-961-083-102 Sequence 102, App
30	525.5	20.7	110	4	US-09-536-784-102 Sequence 102, App
31	499.5	19.7	695	1	US-08-127-499A-23 Sequence 23, Appl
32	499.5	19.7	695	1	US-08-482-847-23 Sequence 23, Appl
33	491	19.4	619	1	US-08-465-746-2 Sequence 2, Appl
34	491	19.4	619	1	US-08-214-164-2 Sequence 2, Appl
35	491	19.4	619	2	US-08-467-852A-3 Sequence 3, Appl
36	491	19.4	619	2	US-08-246-616-2 Sequence 2, Appl
37	491	19.4	619	2	US-08-247-491A-3 Sequence 3, Appl
38	491	19.4	619	2	US-08-319-795-2 Sequence 2, Appl
39	491	19.4	619	2	US-08-468-985-2 Sequence 2, Appl
40	491	19.4	619	3	US-08-312-949-2 Sequence 2, Appl
41	491	19.4	648	1	US-08-072-070-2 Sequence 2, Appl
42	491	19.4	648	1	US-08-468-434-2 Sequence 2, Appl
43	491	19.4	648	1	US-08-214-222-2 Sequence 2, Appl
44	491	19.4	648	2	US-08-467-852A-2 Sequence 2, Appl
45	491	19.4	648	2	US-08-468-718-2 Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-09-286-981B-9
Sequence 9, Application US/09286981B
Patent No. 6503511
GENERAL INFORMATION:
APPLICANT: Wizemann, Theresa M.
APPLICANT: Koenig, Scott
APPLICANT: Johnson, Leslie S
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
FILE REFERENCE: 469201-396
CURRENT FILING DATE: 1999-04-06
PRIOR APPLICATION NUMBER: US 60/085,743
PRIOR FILING DATE: 1998-05-15
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 9
LENGTH: 446
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-286-981B-9

Query Match 86.5%; Score 2194.5; DB 4; Length 446;
Best Local Similarity 99.8%; Pred. No. 7.3e-146;
Matches 445; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY	37	TENEGATQVPTSSNRANESQAEQGHOPKLDSEDPKAEVEEYVKTI VGSYAKTKR	96
DB	1	TENEGATQVPTSSNRANESQAEQGHOPKLDSEDPKAEVEEYVKTI VGSYAKTKR	60
QY	97	HTITATVNEINNIKNEVINKI VESTSESQOILMMESRSKYDEAVSFEKQSSSSSSD	156
DB	61	HTITATVNEINNIKNEVINKI VESTSESQOILMMESRSKYDEAVSFEKQSSSSSSD	120
QY	157	SSTPEASDTAKPKNTEPEGEKVAAEAKKVEEA-KKAKQDEEDRRNPTITTYKTLEI	215
DB	121	SSTPEASDTAKPKNTEPEGEKVAAEAKKVEEA-KKAKQDEEDRRNPTITTYKTLEI	180
QY	216	AESDVEVKKALELVKVVANPEDEOKIKOAEVYESQOAEATRLKIKTREAEEAK	275
DB	181	AESDVEVKKALELVKVVANPEDEOKIKOAEVYESQOAEATRLKIKTREAEEAK	240
QY	276	RRADAKQEGKPKRAKGVPEGLATPPDKENDAKSSSVGEETLPSPSLKPEKKVAEAE	335
DB	241	RRADAKQEGKPKRAKGVPEGLATPPDKENDAKSSSVGEETLPSPSLKPEKKVAEAE	300
QY	336	KKVEBAKKAEQDEEDRRNPTITTYKTLEI AESSDVEVKKALELVKKEAKPRNEEK	395
DB	301	KKVEBAKKAEQDEEDRRNPTITTYKTLEI AESSDVEVKKALELVKKEAKPRNEEK	360

Query 396 VKQAKAEVESKKAELATLEKIKTDKRAKAEBAKKAEDVKVEKPAEOPAPAPAKAEK 455
Db 361 VKQAKAEVESKKAELATLEKIKTDKRAKAEBAKKAEDVKVEKPAEOPAPAPAKAEK 420
Query 456 PAPAPKRENPAPAEOPKAEKPADQQAEE 481
Db 421 PAPAPKRENPAPAEOPKAEKPADQQAEE 446

RESULT 2
US-09-286-981B-6
; Sequence 6, Application US/09286981B
; Patent No. 6503511
; GENERAL INFORMATION:
; APPLICANT: Wizenmann, Theresa M.
; APPLICANT: Koenig, Scott
; APPLICANT: Johnson, Leslie S
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
; PRIORITY REFERENCE: 469201-396
; CURRENT APPLICATION NUMBER: US/09/286,981B
; PRIOR FILING DATE: 1999-04-06
; PRIOR APPLICATION NUMBER: US 60/085,743
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-286-981B-6

Query Match 85.8%; Score 2177.5; DB 4; Length 446;
Best Local Similarity 98.9%; Pred. No. 1,1e-144;
Matches 441; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

Query 37 TENEGATQVPTSSNRANESQAEQGEOPKLDSEBDKARKKEVEEVKKIVGESYAKSTKKR 96
Db 1 TENEGATQVPTSSNRANESQAEQGEOPKLDSEBDKARKKEVEEVKKIVGESYAKSTKKR 60
Query 97 HTTVALVNLNNIKNEYLNKIVESTSESOLQILMESRSKVDVAVSKEFKDSSSSSSSD 156
Db 61 HTTVALVNLNNIKNEYLNKIVESTSESOLQILMESRSKVDVAVSKEFKDSSSSSSSD 120
Query 157 SSTKPEASDPAKNNKPTPEGEKVAEAKKVEEVEKAKDQKEDRRNYPTTYKTLELEI 215
Db 121 SSTKPEASDPAKNNKPTPEGEKVAEAKKVEEVEKAKDQKEDRRNYPTTYKTLELEI 180
Query 216 AESDVEVKKAELVVKVKAEPDEOKIKQAEAEVESKQAEATRLKIKITDREAEBEAK 275
Db 181 AESDVEVKKAELVVKVKAEPDEOKIKQAEAEVESKQAEATRLKIKITDREAEBEAK 240
Query 276 RRADAEQCKPKGRKRGVGEIATPDKKNDAKSSDSSVGEETLPSPLKPEKKVAEAE 335
Db 241 RRADAEQCKPKGRKRGVGEIATPDKKNDAKSSDSSVGEETLPSPLKPEKKVAEAE 300
Query 336 KKYVEAKKKAEDQKEDRRNYPTNTYKTLELEIAESVVEVKKAELVKEAEKPEPNEEK 395
Db 301 KKYVEAKKKAEDQKEDRRNYPTNTYKTLELEIAESVVEVKKAELVKEAEKPEPNEEK 360
Query 396 VKQAKAEVESKKAELATLEKIKTDKRAKAEBAKKAEDVKVEKPAEOPAPAPAKAEK 455
Db 361 VKQAKAEVESKKAELATLEKIKTDKRAKAEBAKKAEDVKVEKPAEOPAPAPAKAEK 420
Query 456 PAPAPKRENPAPAEOPKAEKPADQQAEE 481
Db 421 PAPAPKRENPAPAEOPKAEKPADQQAEE 446

RESULT 3
US-09-308-022-6
; Sequence 6, Application US/09308022
; Patent No. 6291654
; GENERAL INFORMATION:

APPLICANT: REGENTS OF THE UNIVERSITY OF MINNESOTA, ET AL.
TITLE OF INVENTION: C3 BINDING PROTEIN OF STREPTOCOCCUS
TITLE OF INVENTION: PNEUMONIAE
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEETING, RAASCH & GEBHARDT, P.A.
STREET: 119 No. 6291654th Fourth Street, Suite 203
CITY: Minneapolis
STATE: Minnesota
COUNTRY: USA
ZIP: 55401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/308,022
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/20586
FILING DATE: 12-NOV-1997
APPLICATION NUMBER: 60/029,444
FILING DATE: 12-NOV-1996
APPLICATION NUMBER: 60/038,086
FILING DATE: 18-FEB-1997
APPLICATION NUMBER: 60/059,368
FILING DATE: 19-SEP-1997
APPLICATION NUMBER: 60/062,473
FILING DATE: 16-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: MOETING, Ann M.
REGISTRATION NUMBER: 33,977
REFERENCE/DOCKET NUMBER: 110 00430101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-305-1217
TELEFAX: 612-305-1228
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 564 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-308-022-6

Query Match 68.1%; Score 1728; DB 3; Length 564;
Best Local Similarity 70.5%; Pred. No. 3.8e-113;
Matches 370; Conservative 50; Mismatches 47; Indels 58; Gaps 10;

Query 1 MPASKSEKRVHYSIRKFSVG-ASVVVASLWGSVVHATENEGATQVPTSSNRANESQAEQ 59
Db 1 MPASKSEKRVHYSIRKFSVGASVVVASLWGSVVHATENEGATQVPTSSNRANESQAEQ 54
Query 60 GEQPKLDSEBDKARKKE-VEEVKKIVGESYAKSTKKRTTIVAVLNNIKNEYLNK- 117
Db 55 -----TEHRKAKQVVDYIEKRLR--IOLDRKRTQVAVLNNIKNEYLNK- 104
Query 118 -IVESTSESOLQILMESRSKVDVAVSKEFKDSSSSSSDSTKPEASDPAKNNKPTPEB 176
Db 105 NVLEESKQDEL-SEIRAKIDAAFEKTKQT-----LKG 138
Query 177 EKVAAEKKKVVEA-KKAKDQKEDRRNYPTTYKTLELEIAESVVEVKKAELVVKVKA 235
Db 139 EKVAAEKKKVVEA-KKAKDQKEDRRNYPTNTYKTLELEIAEFVVKAELELVKAEK 198
Query 236 EPRDEOKIKQAEAEVESKQAEATRLKIKITDREAEBEAKRADAK-----EQGRP 286
Db 199 ESRNBTGTTQAKAEKESKKAELATLEKIKTDKRAKAEBAKKAEDVKVEKPAEOPAPAKAEK 258
Query 287 KGRKRGVGEIATPDKKNDAKSSDSSVGEETLPSPLKPEKKVAEAKKVEAEKKAAB 346

Db 259 KGRKRGVPGELATPDKKENDAKSSDSSVGEETLPSSSLKSGKVAEAKKEVEAEKKA 318
Qy 347 DQKEDRRNPTNTYKTLLEIAESDVVKAELELVKEAKPRNEEKKQAKAEVSK 406
Db 319 DQKEDRRNPTNTYKTLLEIAESDVVKAELELVKEAKPRNEEKKQAKAEVSK 378
Qy 407 KAEATRLKIKITDRKKAEEBAKKAEDVKKEPAEOPAPAPAKAEKPAKPEKPA 466
Db 379 KAEATRLKIKITDRKKAEEBAKKAEDVKKEPAEOPAPAPAKAEKPAKPEKPA 436
Qy 467 EQPKAEKPADQAEEDYARSRSEERYNLTQOOPKTEKPAOPSTP 511
Db 437 EQPKAEKPADQAEEDYARSRSEERYNLTQOOPKTEKPAOPSTP 481

RESULT 4
US-09-286-981B-18
Sequence 18, Application US/09286981B
Patent No. 6503511
GENERAL INFORMATION:
APPLICANT: Koenig, Scott
APPLICANT: Johnson, Leslie S
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
FILE REFERENCE: 469201-396
CURRENT APPLICATION NUMBER: US/09/286,981B
CURRENT FILING DATE: 1999-04-06
PRIOR APPLICATION NUMBER: US 60/085,743
PRIOR FILING DATE: 1998-05-15
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 18
LENGTH: 406
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-286-981B-18

Query Match 59.6%; Score 1512.5; DB 4; Length 406;
Best Local Similarity 73.0%; Pred. No. 2.9e-98;

Matches 327; Conservative 28; Mismatches 46; Indels 47; Gaps 8;

Qy 37 TENEGATOVPTSSNRANESQAEQEQPKLDSERDARKEVEEYKIKVGEASYAKSTKR 96
Db 1 TENEGTTOAPTSSNRGNEQA-----EHMKAAQVDYIEKML-----QDDRK 44
Qy 97 HTTVLALVNLNNIKKYEYLNKIVESTESQLOLIMMESKSKVDNAVSKTEKSSSSSSD 156
Db 45 HTQNVGLTLTGAIKTEYLRGLSVSKKSTYAE-LPSEIKKLTAAFKQPKK----- 94
Qy 157 SSTKPEASDTAKPNKPTPEGEKVAEAKKVEEA-KKAKQKEDDRNYPPTTYKTLLEI 215
Db 95 -----DTLAKK-----KVAEAKKVAEAKKAEEDQKEDDRNYPPTTYKTLLEI 140
Qy 216 AESDVVKAELELVVKANEPEDQIKQAEAEVSKQAEATRLKKITDRBEAEK 275
Db 141 AESDVVKAELELVVKANEPEDQIKQAEAEVSKQAEATRLKKITDRBEAEK 200
Qy 276 RRDADAEQ-----KPKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPSGLPEKKV 331
Db 201 RRDADAEQ-----KPKGRKRGVPGELATPDKKENDAKSSDSSVGEETLPSGLPEKKV 260
Qy 332 AEAKEKVEAEKKAEDQKEDDRNYPPTTYKTLLEIAESDVVKAELELVKEAKEPR 391
Db 261 AEAKEKVEAEKKAEDQKEDDRNYPPTTYKTLLEIAESDVVKAELELVKEAKEPR 320
Qy 392 NEKVQKAAVEVSKQAEATRLKIKITDRKKAEEBAKKAEDVKKEPAEOPAPAP 451
Db 321 NEKVQKAAVEVSKQAEATRLKIKITDRKKAEEBAKKAEDVKKEPAEOPAPAP 380
Qy 452 KAEKPAKPEKPAEOPAPAEKPADQQA 479
Db 381 QPK-----PAKPEKPAEOPAPAEKPADQQA 406

RESULT 5
US-08-847-065-25
Sequence 25, Application US/08847065
Patent No. 6245335

GENERAL INFORMATION:
APPLICANT: Masure, H. Robert
APPLICANT: Rosenow, Carsten I.
APPLICANT: Tuomanen, Elaine
APPLICANT: Witzmann, Theresa M.
TITLE OF INVENTION: CHOLINE BINDING PROTEINS FOR
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: David A. Jackson, Esq.
STREET: 411 Hackensack Ave, Continental Plaza, 4th
STREET: Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/847,065
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-158
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-1684
TELEFAX: 201-343-1684

INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 631 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
FRAGMENT TYPE: N-terminal
US-08-847-065-25

Query Match 59.4%; Score 1506; DB 3; Length 631;
Best Local Similarity 73.5%; Pred. No. 1.4e-97;

Matches 316; Conservative 39; Mismatches 35; Indels 40; Gaps 6;

Qy 94 KKRHTTVLALVNLNNIKKYEYLNK--IVESTESQLOLIMMESKSKVDNAVSKTEKSSS 151
Db 9 RKRHTQVVALNLIKSAITKTYLRLNVLVEKSKDEL--SEIKAKLAAFEKFKDT-- 63
Qy 152 SSSDSSSTKPEASDTAKPNKPTPEGEKVAEAKKVEEA-KKAKQKEDDRNYPPTTYKT 210
Db 64 -----LKPGEKVAEAKKVEEA-KKAEEDQKEDDRNYPPTTYKT 102
Qy 211 LELEIAESDVVKAELELVVKANEPEDQIKQAEAEVSKQAEATRLKKITDRBEA 270
Db 103 LELEIAESDVVKAELELVVKANEPEDQIKQAEAEVSKQAEATRLKKITDRKA 162
Qy 271 EEAARRADAK-----EQPKGRKRGVPGELATPDKKENDAKSSDSSVGEETLP 321
Db 163 EEAARRADAK-----EQPKGRKRGVPGELATPDKKENDAKSSDSSVGEETLP 222
Qy 322 SPGLPEKKVVAEAKKVEAEKKAEDQKEDDRNYPPTTYKTLLEIAESDVVKAELE 381
Db 223 SSSLSKSGKVAEAKKVEAEKKAEDQKEDDRNYPPTTYKTLLEIAESDVVKAELE 282
Qy 382 LVKEAKEPRNEEYKQAKAEVSKQAEATRLKIKITDRKKAEEBAKKAEDVKKEKPA 441

Db 283 LVEEAEPEDEEETIKQAKAKVESKKAETRLNKTDRKKAEEBAKKAEDVKAEKP 342
Qy 442 AEOPAPAPAKAEKAPAPENPAEPKAEKPADQOAEEDVYARSEENRLLTQOQPPK 501
Db 343 AEOPAPAPAPQEPK-PAPPEKPAEOPKAEKTDQOAEEDVYARSEENRLLTQOQPPK 400
Qy 502 TEKPAOPSTP 511
Db 401 TEKPAOPSTP 410

RESULT 6
US-09-286-981B-12
; Sequence 12, Application US/09286981B
; Patent No. 6503511
; GENERAL INFORMATION:
; APPLICANT: Wizemann, Theresa M.
; APPLICANT: Koenig, Scott
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
; FILE REFERENCE: 469201-396
; CURRENT APPLICATION NUMBER: US/09/286,981B
; CURRENT FILING DATE: 1999-04-06
; PRIOR APPLICATION NUMBER: US 60/085,743
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 12
; LENGTH: 426
; TYPE: PR1
; ORGANISM: Streptococcus pneumoniae
US-09-286-981B-12

Query Match 57.4%; Score 1455.5; DB 4; Length 426;
Best Local Similarity 68.4%; Pred. No. 3e-94;
Matches 318; Conservative 41; Mismatches 47; Indels 59; Gaps 9;

Qy 37 TENEGATQVPTSSNRANESQAEQEPKLDSEEDKARKEVEEYKIVGESYAKSTKR 96
Db 1 TEKEVTTQVATSSNRANKSQ-----TEHMKAKQVDEYIKKKI-----QIDRRK 44
Qy 97 HTTVALVNLNNKXNKLIVESTESQOILMESRSKYDEAVSKFEKSSSSSSD 156
Db 45 HTQNVGLTLKGLVITKTEYHLGSLVSKKSEAE-LPSEIKALIDAPQPKDT----- 96
Qy 157 SSTPEASDTAKPNKPTPEGEKVAEAKKVEEA-KKAKDOKEEDRRNYPTIYKTLLEI 215
Db 97 -----LPTEPGKVAEAKKVEEAKKKAEDQEKDRLNRYPTNYKTLLELDI 142
Qy 216 AESDVEYKAELELVYKANEPRDEOKIKQAEAVESKQAEATRLKIKITDREAEBAK 275
Db 143 AESDVEYKAELELVKEEAKESRDEKINQAKAVENKKAETRLKIKITDREKA-BEAK 201
Qy 276 RRADAK-----EQGPKGRAGVPGELATPDKKENDAKSSDSSVGEETLTPSLK 326
Db 202 RRADAKQAEANVATSEODKSRRAKREVLGELATPDKKENDAKSSDSSVGEETLTPSLK 261
Qy 327 PEKKVAEAEKKEEAKKKAEDQEKEDRRNYPTNYKTLLEIAESDVEYKAELELVKEE 386
Db 262 PEKKVAEAEKKEEAKKKAEDQEKEDRRNYPTNYKTLLEIAESDVEYKAELELVKEE 321
Qy 387 AKEPRNEKYKQAEVESKKAETRLKIKITDRKKA-BEBAKKAABEDKVEKPAEOP 445
Db 322 AKESRNEEKIKQVAKVESKKAETRLKIKITDRKKAEEBAKRAAEEDKVEKPAEOP 381
Qy 446 QPAPAPAKAEK-----PAPAPKPNPAOPKAEKPADQOAE 481
Db 382 QPAPAPQEPKTEEPENPAAPAPKPNPAEKPAEKPADQOAE 426

RESULT 7
US-09-286-981B-13

; Sequence 13, Application US/09286981B
; Patent No. 6503511
; GENERAL INFORMATION:
; APPLICANT: Wizemann, Theresa M.
; APPLICANT: Koenig, Scott
; APPLICANT: Koenig, Scott
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
; FILE REFERENCE: 469201-396
; CURRENT APPLICATION NUMBER: US/09/286,981B
; CURRENT FILING DATE: 1999-04-06
; PRIOR APPLICATION NUMBER: US 60/085,743
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 13
; LENGTH: 425
; TYPE: PR1
; ORGANISM: Streptococcus pneumoniae
US-09-286-981B-13

Query Match 57.1%; Score 1447.5; DB 4; Length 425;
Best Local Similarity 68.3%; Pred. No. 1.1e-93;
Matches 317; Conservative 41; Mismatches 47; Indels 59; Gaps 9;

Qy 37 TENEGATQVPTSSNRANESQAEQEPKLDSEEDKARKEVEEYKIVGESYAKSTKR 96
Db 1 TEKEVTTQVATSSNRANKSQ-----TEHMKAKQVDEYIKKKI-----QIDRRK 44
Qy 97 HTTVALVNLNNKXNKLIVESTESQOILMESRSKYDEAVSKFEKSSSSSSD 156
Db 45 HTQNVGLTLKGLVITKTEYHLGSLVSKKSEAE-LPSEIKALIDAPQPKDT----- 96
Qy 157 SSTPEASDTAKPNKPTPEGEKVAEAKKVEEA-KKAKDOKEEDRRNYPTIYKTLLEI 215
Db 97 -----LPTEPGKVAEAKKVEEAKKKAEDQEKDRLNRYPTNYKTLLELDI 142
Qy 216 AESDVEYKAELELVYKANEPRDEOKIKQAEAVESKQAEATRLKIKITDREAEBAK 275
Db 143 AESDVEYKAELELVKEEAKESRDEKINQAKAVENKKAETRLKIKITDREKA-BEAK 201
Qy 276 RRADAK-----EQGPKGRAGVPGELATPDKKENDAKSSDSSVGEETLTPSLK 326
Db 202 RRADAKQAEANVATSEODKSRRAKREVLGELATPDKKENDAKSSDSSVGEETLTPSLK 261
Qy 327 PEKKVAEAEKKEEAKKKAEDQEKEDRRNYPTNYKTLLEIAESDVEYKAELELVKEE 386
Db 262 PEKKVAEAEKKEEAKKKAEDQEKEDRRNYPTNYKTLLEIAESDVEYKAELELVKEE 321
Qy 387 AKEPRNEKYKQAEVESKKAETRLKIKITDRKKA-BEBAKKAABEDKVEKPAEOP 445
Db 322 AKESRNEEKIKQVAKVESKKAETRLKIKITDRKKAEEBAKRAAEEDKVEKPAEOP 381
Qy 446 QPAPAPAKAEK-----PAPAPKPNPAOPKAEKPADQOAE 480
Db 382 QPAPAPQEPKTEEPENPAAPAPKPNPAEKPAEKPADQOAE 425

RESULT 8
US-09-286-981B-14
; Sequence 14, Application US/09286981B
; Patent No. 6503511
; GENERAL INFORMATION:
; APPLICANT: Wizemann, Theresa M.
; APPLICANT: Koenig, Scott
; APPLICANT: Koenig, Scott
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
; FILE REFERENCE: 469201-396
; CURRENT APPLICATION NUMBER: US/09/286,981B
; CURRENT FILING DATE: 1999-04-06
; PRIOR APPLICATION NUMBER: US 60/085,743
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 14
LENGTH: 424
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-286-981B-14

Query Match 57.0%; Score 1446.5; DB 4; Length 424;
Best Local Similarity 68.5%; Pred. No. 1.2e-93;
Matches 317; Conservative 40; Mismatches 47; Indels 59; Gaps 9;

QY 37 TENEGATQVPTSSNRANESQAEQEPKLDSEPDKARKEVEEYVKKIIGESYAKTKKR 96
DB 1 TEKEVTTQVATSSNRANESQAEQEPKLDSEPDKARKEVEEYVKKIIGESYAKTKKR 44
QY 97 HTITVALVNEINNIKNKYLNKIVESTSESQOLIMMESRSGVDEAVSKPEKSSSSSSD 156
DB 45 HTQVGLTTLGVITKTEYLHGLSVSKKSEAE-LPSEIKAKLDAAEFQFKDT----- 96
QY 157 SSTKPEASDTAKPNKPTPEGEKVAEAKKVEEA-KKAKDQKEDRRNYPITTYKLELEI 215
DB 97 -----LPTEPGKVAEAKKVEEA-KKAKDQKEDRRNYPITTYKLELEI 142
QY 216 AESDVEVKKAELEIVKVAKEPRDEOKIKQAEAVESKQAEATRLKKIKTDREAEAEAK 275
DB 143 AESDVEVKKAELEIVKVAKEPRDEOKIKQAEAVESKQAEATRLKKIKTDREAEAEAK 201
QY 276 RRADAK-----EGGPKGRAGVGPGLATPDKKENDAKSSDSSVGEETLPSLSLK 326
DB 202 RRADAKQOEANVATSEODKSRKRAKREVLDELATPDKKENDAKSSDSSVGEETLPSLSLK 261
QY 327 PEKKVAEAKKVEEAKKAEADQKEDRRNYPITTYKLELEIAESDVEVKKAELEIVKEE 386
DB 262 PEKKVAEAKKVEEAKKAEADQKEDRRNYPITTYKLELEIAESDVEVKKAELEIVKEE 321
QY 387 AKERNEEYKQAKAEVSKAEATRLKKIKTDREAEAEAK-EEBAKRAAEEDKVEKPAEOP 445
DB 322 AKSRNEEKIKQYAKAEVSKAEATRLKKIKTDREAEAEAK-EEBAKRAAEEDKVEKPAEOP 381
QY 446 QPAPAPAEK-----PAPAPKPEPAPAEOPKAEKPAEOP 479
DB 382 QPAPAPAEKPEPAPAEOPKAEKPAEOP 424

RESULT 9

US-09-286-981B-7
Sequence 7, Application US/09286981B
Patent No. 6503511
GENERAL INFORMATION:
APPLICANT: Wizemann, Theresa M.
PLICANT: Koenig, Scott
PLICANT: Johnson, Leslie S
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
FILE REFERENCE: 469201-396
CURRENT APPLICATION NUMBER: US/09/286,981B
PRIOR FILING DATE: 1999-04-06
PRIOR APPLICATION NUMBER: US 60/085,743
PRIOR FILING DATE: 1998-05-15
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Patent Ver. 2.1
SEQ ID NO 7
LENGTH: 428
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-286-981B-7

Query Match 56.8%; Score 1440.5; DB 4; Length 428;
Best Local Similarity 69.7%; Pred. No. 3.3e-93;
Matches 311; Conservative 35; Mismatches 53; Indels 47; Gaps 7;

QY 60 GEGPKLDEBDKAR---EVEEYVKKIIGESYAKTKKRHTITVALVNEINNIKNKYLN 116
DB 6 GNNSTVTSGGODISKYADEVESHQSLIKDVNKLKKTQHTQNDPNNKSLKITKTKYLY 65

QY 117 K-IVESTESOLQILIMMESRSGVDEAVSKPEKSSSSSSDSTKPEASDTAKPNKPT 174
DB 66 ELNVLEKSEBAELTSTKTEKTELTAAEFQFKDTLS-----IT 104
QY 175 PGEKVAEAKKVEEA-KKAKDQKEDRRNYPITTYKLELEIAESDVEVKKAELEIVK 233
DB 105 PEKKVAEAKKVEEAKKAEADQKEDRRNYPITTYKLELEIAESDVEVKKAELEIVK 164
QY 234 ANPRDEOKIKQAEAVESKQAEATRLKKIKTDREAE-----EBAKRAEAE 282
DB 165 ANPRDEOKIKQAEAVESKQAEATRLKKIKTDREAEATRLKKIKTDREAEAEAEK 224
QY 283 OGPKGRAGVGPGLATPDKKENDAKSSDSSVGEETLPSLSLKPEKVAEAKKVEEAK 342
DB 225 E-PKRTKRGVLGEATPDKKENDAKSSDSSVGEETLPSLSLKPEKVAEAKKVEEAK 282
QY 343 KAEADQKEDRRNYPITTYKLELEIAESDVEVKKAELEIVKEAEPRNEEKYQAKAE 402
DB 283 KAEADQKEDRRNYPITTYKLELEIAESDVEVKKAELEIVKEAEPRNEEKYQAKAE 342
QY 403 VESKQAEATRLKKIKTDREAEAEAKKAEADKVEKPAEOPAE-----PAPAEK 455
DB 343 VESKQAEATRLKKIKTDREAEAEAKKAEADKVEKPAEOPAE-----PAPAEK 402
QY 456 PAPAPKPEPAPAEOPKAEKPAEOP 481
DB 403 PAPAPKPEPAPAEOPKAEKPAEOP 428

RESULT 10

US-09-286-981B-11
Sequence 11, Application US/09286981B
Patent No. 6503511
GENERAL INFORMATION:
APPLICANT: Wizemann, Theresa M.
PLICANT: Koenig, Scott
PLICANT: Johnson, Leslie S
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
FILE REFERENCE: 469201-396
CURRENT APPLICATION NUMBER: US/09/286,981B
PRIOR FILING DATE: 1999-04-06
PRIOR APPLICATION NUMBER: US 60/085,743
PRIOR FILING DATE: 1998-05-15
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Patent Ver. 2.1
SEQ ID NO 11
LENGTH: 425
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-286-981B-11

Query Match 56.4%; Score 1432; DB 4; Length 425;
Best Local Similarity 66.6%; Pred. No. 1.3e-92;
Matches 311; Conservative 44; Mismatches 48; Indels 64; Gaps 9;

QY 37 TENEGATQVPTSSNRANESQAEQEPKLDSEPDKARKEVEEYVKKIIGESYAKTKKR 96
DB 1 TEKEVTTQVATSSNRANESQAEQEPKLDSEPDKARKEVEEYVKKIIGESYAKTKKR 44
QY 97 HTITVALVNEINNIKNKYLNK-IVESTESOLQILIMMESRSGVDEAVSKPEKSSSSSS 154
DB 45 HTQVGLTTLGVITKTEYLHGLSVSKKSEAE-LPSEIKAKLDAAEFQFKDTNRITK 101
QY 155 SDSSTKPEASDTAKPNKPTPEGEKVAEAKKVEEA-KKAKDQKEDRRNYPITTYKLEL 213
DB 102 T-----VAEAEKVEEAKKAEADQKEDRRNYPITTYKLEL 138
QY 214 EIASDVEVKKAELEIVKVAKEPRDEOKIKQAEAVESKQAEATRLKKIKTDREAEAE 273
DB 139 EIASDVEVKKAELEIVKVAKEPRDEOKIKQAEAVESKQAEATRLKKIKTDREAEAE 198
QY 274 AKRPAEAK-----EGGPKGRAGVGPGLATPDKKENDAKSSDSSVGEETLPS 321

PRIOR FILING DATE: 1998-05-15
 NUMBER OF SEQ ID NOS: 38
 SOFTWARE: Patent In Ver. 2.1
 SEQ ID NO 16
 LENGTH: 414
 TYPE: PRT
 ORGANISM: Streptococcus pneumoniae
 US-09-286-981B-16

Query Match 55.4%; Score 1405.5; DB 4; Length 414;
 Best Local Similarity 67.7%; Pred. No. 8,9e-91;
 Matches 310; Conservative 45; Mismatches 46; Indels 57; Gaps 11;

QY TENEGATQVPTSSNRANESQAEQGEQPKLDSERDKARKEVEEYVKIIGSVAKSTKK 95
 1 TENEGSTQATSSNMAK-----TEHRAAKQVVDIEYIKMLRE--IQLDNR 44
 Db
 QY 96 RHITVALVNLNNIKNEIYLNK--IVESTSESQIOLIMESRSKVDKAVSKFEKSSSS 153
 45 KHTQVVALNLIKLSAIKTKYLRLEINVLSEKSKDEL P--SEIKAKLDAA--FEK----- 92
 Db
 QY 154 SSSSTKPEASDTAKPNKPTPEGEKVAEAKKVEEA-KKAKDQKEEDRRNYPPTTYKTLE 212
 93 -----EKKDTLK-----PGEKVAEAKKVEEAKKADQKEEDRRNYPPTTYKTLE 138
 Db
 QY 213 LEIASDVEYKKALELVKYNANPRDEOKIKQAEAVESKQAEATRLKIKITDRBEAE 272
 139 LEIAEDVVKKAELELVKAEKESNNEGITKQAEKVESKKAETRLNLIKTRDKAAE 198
 Db
 QY 273 EAKRADAK-----EQQKPKGAKGVPELATPDKKENDAKSSSVGEETLPS 323
 199 EAKRKADAKLKEANVATSDQKPKGAKGVPELATPDKKENDAKSSSVGEETLPS 258
 Db
 QY 324 SLKPEKKVAEAKKVEEAKKADQKEEDRRNYPPTTYKTLELEIASDVEYKKALELV 383
 259 SLKSGKVAEAKKVEEAKKADQKEEDRRNYPPTTYKTLELEIASDVEYKKALELV 318
 Db
 QY 384 KEAKEPRNEEKVQAKAEVSKKAEATRLKIKTRKKAEEAKKAAEEDVKKEKPAE 443
 319 KEAKEPRDEEIKQAKAKVESKKAETRLNLIKTRKQAEKAAEEDVKKEKPAE 378
 Db
 QY 444 QOPAPAPAKPAKPAKPEPNPAEQPAEKPADQAAE 481
 379 QOPAPATQPEK--PAPKPEKPAEQPAEKTDQAAE 414
 Db

RESULT 14
 US-09-286-981B-10
 Sequence 10, Application US/09286981B
 Patent No. 6503511
 GENERAL INFORMATION:
 APPLICANT: Wismann, Theresa M.
 APPLICANT: Koenig, Scott
 APPLICANT: Johnson, Leslie S
 TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
 FILE REFERENCE: 469201-396
 CURRENT APPLICATION NUMBER: US/09/286,981B
 PRIOR FILING DATE: 1998-05-15
 NUMBER OF SEQ ID NOS: 38
 SOFTWARE: Patent In Ver. 2.1
 SEQ ID NO 10
 LENGTH: 414
 TYPE: PRT
 ORGANISM: Streptococcus pneumoniae
 US-09-286-981B-10

Query Match 55.2%; Score 1401.5; DB 4; Length 414;
 Best Local Similarity 66.6%; Pred. No. 1.7e-90;
 Matches 305; Conservative 48; Mismatches 48; Indels 57; Gaps 9;

QY 37 TENEGATQVPTSSNRANESQAEQGEQPKLDSERDKARKEVEEYVKIIGSVAKSTKK 95

Db
 1 TENEGSTQATSSNMAK-----TEHRAAKQVVDIEYIKMLRE--IQLDNR 44
 QY 96 RHITVALVNLNNIKNEIYLNK--IVESTSESQIOLIMESRSKVDKAVSKFEKSSSS 153
 45 KHTQVVALNLIKLSAIKTKYLRLEINVLSEKSKDEL P--SEIKAKLDAAFEKFKOT---- 97
 Db
 QY 154 SSSSTKPEASDTAKPNKPTPEGEKVAEAKKVEEA-KKAKDQKEEDRRNYPPTTYKTLE 212
 98 -----LKPGEKVAEAKKVEEAKKADQKEEDRRNYPPTTYKTLE 138
 Db
 QY 213 LEIASDVEYKKALELVKYNANPRDEOKIKQAEAVESKQAEATRLKIKITDRBEAE 272
 139 LEIAEDVVKKAELELVKAEKESNNEGITKQAEKVESKKAETRLNLIKTRDKAAE 198
 Db
 QY 273 EAKRADAK-----EQQKPKGAKGVPELATPDKKENDAKSSSVGEETLPS 323
 199 EAKRKADAKLKEANVATSDQKPKGAKGVPELATPDKKENDAKSSSVGEETLPS 258
 Db
 QY 324 SLKPEKKVAEAKKVEEAKKADQKEEDRRNYPPTTYKTLELEIASDVEYKKALELV 383
 259 SLKSGKVAEAKKVEEAKKADQKEEDRRNYPPTTYKTLELEIASDVEYKKALELV 318
 Db
 QY 384 KEAKEPRNEEKVQAKAEVSKKAEATRLKIKTRKKAEEAKKAAEEDVKKEKPAE 443
 319 KEAKEPRDEEIKQAKAKVESKKAETRLNLIKTRDRAEEAKKAAEEDVKKEKPAE 378
 Db
 QY 444 QOPAPAPAKPAKPAKPEPNPAEQPAEKPADQAAE 481
 379 QOPAPATQPEK--PAPKPEKPAEQPAEKTDQAAE 414
 Db

RESULT 15
 US-09-286-981B-3
 Sequence 3, Application US/09286981B
 Patent No. 6503511
 GENERAL INFORMATION:
 APPLICANT: Wismann, Theresa M.
 APPLICANT: Koenig, Scott
 APPLICANT: Johnson, Leslie S
 TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
 FILE REFERENCE: 469201-396
 CURRENT APPLICATION NUMBER: US/09/286,981B
 PRIOR FILING DATE: 1998-05-15
 NUMBER OF SEQ ID NOS: 38
 SOFTWARE: Patent In Ver. 2.1
 SEQ ID NO 3
 LENGTH: 431
 TYPE: PRT
 ORGANISM: Streptococcus pneumoniae
 US-09-286-981B-3

Query Match 55.0%; Score 1395; DB 4; Length 431;
 Best Local Similarity 64.0%; Pred. No. 5.1e-90;
 Matches 304; Conservative 57; Mismatches 60; Indels 54; Gaps 9;

QY 37 TENEGATQVPTSSNRANESQAEQGEQPKLDSERDKARKEVEEYVKIIGSVAKSTKK 96
 1 TEKEVTTVATSSNMANKSQ-----TEHRAAKQVVDIEYIKMLRE--IQLDNR 44
 Db
 QY 97 HTITVALVNLNNIKNEIYLNK--IVESTSESQIOLIMESRSKVDKAVSKFEKSSSSSD 156
 45 HTQVVALNLIKLSAIKTKYLRLEINVLSEKSKDELSTKKEIDAFQEFNDT----- 97
 Db
 QY 157 SSTKPEASDTAKPNKPTPEGEKVAEAKKVEEA-KKAKDQKEEDRRNYPPTTYKTLEI 215
 98 -----LKPGEKVAEAKKVEEAKKADQKEEDRRNYPPTTYKTLEI 141
 Db
 QY 216 AEDSVEYKKALELVKYNANPRDEOKIKQAEAVESKQAEATRLKIKITDRBEAEBAK 275
 142 AEDSVEYKKALELVKAEKESNNEGITKQAEKVESKKAETRLNLIKTRDKAAEBAK 201
 Db

QY 276 RRAADAKEQ--GKPKGRAKRGVPGELATPDKKENDAKSSDSSVGEETLPSPSLK----- 326
Db 202 RKAEEVEVKNKCKKRTKRGAFGEPA TPDKKENDAKSSDSSVVKKS-SKPILKSEKKVAE 260
QY 327 PEKKVAEAEKKVEAEAKKEDOKEDERRNYPTNTYKTLELEIAESDVEVKKAELVKEE 386
Db 261 AEKKVAEAEKKVAEAEKKAADOKEDERRNYPTNTYKTLELEIAESDVKKAELELVKEE 320
QY 387 AKEPNNEKKVKAQAEVESKKAETRLLEKI KTDKKAEEBAKRRKAAEEDVKVEKPAEQPQ 446
Db 321 AKEPQNEKIKQAKAKVESKKAETRLLEKIKTDKKA-EEAKRKAEEEDVKVEKPAEQPQ 379
QY 447 PAPAPKAEPAPAPKPEPAPAEQPKAEKPADQQAEEEDYARSSSEBYNRLTQOOPPK 501
Db 380 PAPAP--KPAPAPQPEKPAEQPKAEKPADQQAEEEDYARSSSEBYNPLDLTAPAK 431

h completed: November 21, 2003, 13:35:10
ime : 22.4457 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 21, 2003, 13:30:37 ; Search time 7.55427 Seconds
(without alignments)
1008.166 Million cell updates/sec

Title: US-09-298-523C-13_COPY_263_442
Perfect score: 889
Sequence: 1 IKTDREAEAEAKRRADAKE.....AEEBAKRAAEEDKVEKRA 180

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: Issued Patents_AA:*
- 2: /cgn2_6/prodata/1/iaa/5A COMB.pep:*
- 3: /cgn2_6/prodata/1/iaa/6A COMB.pep:*
- 4: /cgn2_6/prodata/1/iaa/6B COMB.pep:*
- 5: /cgn2_6/prodata/1/iaa/PCTUS COMB.pep:*
- 6: /cgn2_6/prodata/1/iaa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	889	100.0	446	4	US-09-286-981B-9 Sequence 9, Appl
2	889	100.0	453	3	US-08-961-083-38 Sequence 38, Appl
3	889	100.0	453	4	US-09-536-784-38 Sequence 38, Appl
4	884	99.4	446	4	US-09-286-981B-6 Sequence 6, Appl
5	803.5	90.4	414	4	US-09-286-981B-16 Sequence 16, Appl
6	803.5	90.4	631	3	US-08-847-065-25 Sequence 25, Appl
7	801.5	90.2	564	3	US-09-308-022-6 Sequence 6, Appl
8	792	89.1	406	4	US-09-286-981B-18 Sequence 18, Appl
9	792	89.1	425	4	US-09-286-981B-11 Sequence 11, Appl
10	791.5	89.0	414	4	US-09-286-981B-10 Sequence 10, Appl
11	784	88.2	428	4	US-09-286-981B-7 Sequence 7, Appl
12	775.5	87.2	412	4	US-09-286-981B-17 Sequence 17, Appl
13	770.5	86.7	424	4	US-09-286-981B-14 Sequence 14, Appl
14	770.5	86.7	426	4	US-09-286-981B-12 Sequence 12, Appl
15	769.5	86.6	425	4	US-09-286-981B-13 Sequence 13, Appl
16	768	86.4	419	4	US-09-286-981B-15 Sequence 15, Appl
17	717.5	80.7	413	4	US-09-286-981B-5 Sequence 5, Appl
18	684.5	77.0	864	4	US-08-714-741-40 Sequence 40, Appl
19	676.5	76.1	431	4	US-09-286-981B-3 Sequence 3, Appl
20	669	75.3	588	4	US-08-714-741-42 Sequence 42, Appl
21	557.5	62.7	219	4	US-09-286-981B-8 Sequence 8, Appl
22	542	61.0	114	4	US-09-286-981B-19 Sequence 19, Appl
23	521	58.6	142	3	US-08-847-065-21 Sequence 21, Appl
24	517	58.2	251	4	US-09-286-981B-4 Sequence 4, Appl
25	487	54.8	103	4	US-09-286-981B-1 Sequence 1, Appl
26	461.5	51.9	605	4	US-08-714-741-46 Sequence 46, Appl
27	451	50.7	103	4	US-09-286-981B-38 Sequence 38, Appl

28	447.5	50.3	1231	4	US-08-714-741-41 Sequence 41, Appl
29	394.5	44.4	623	4	US-08-714-741-47 Sequence 47, Appl
30	375.5	42.2	8991	4	US-08-714-741-32 Sequence 32, Appl
31	306	34.4	110	3	US-08-961-083-102 Sequence 102, Appl
32	306	34.4	110	4	US-09-536-784-102 Sequence 102, Appl
33	237.5	26.7	55	4	US-09-286-981B-37 Sequence 37, Appl
34	182.5	20.5	1507	3	US-08-929-329-5 Sequence 5, Appl
35	174	19.6	468	4	US-09-328-352-6321 Sequence 6321, Ap
36	159.5	17.9	611	4	US-09-216-393B-81 Sequence 81, Appl
37	158.5	17.8	288	3	US-08-312-949-4 Sequence 4, Appl
38	158.5	17.8	288	3	US-08-446-201-4 Sequence 4, Appl
39	158.5	17.8	619	1	US-08-465-746-2 Sequence 2, Appl
40	158.5	17.8	619	1	US-08-214-164-2 Sequence 2, Appl
41	158.5	17.8	619	2	US-08-467-852A-3 Sequence 3, Appl
42	158.5	17.8	619	2	US-08-246-636-2 Sequence 2, Appl
43	158.5	17.8	619	2	US-08-247-491A-3 Sequence 3, Appl
44	158.5	17.8	619	2	US-08-319-795-2 Sequence 2, Appl
45	158.5	17.8	619	2	US-08-468-985-2 Sequence 2, Appl

ALIGNMENTS

```

RESULT 1
US-09-286-981B-9
; Sequence 9, Application US/09286981B
; Patent No. 6503511
; GENERAL INFORMATION:
; APPLICANT: Wizenmann, Theresa M.
; APPLICANT: Koenig, Scott
; APPLICANT: Johnson, Leslie S
; TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
; FILE REFERENCE: 469201-336
; CURRENT APPLICATION NUMBER: US/09/286,981B
; PRIOR APPLICATION NUMBER: 1999-04-06
; PRIOR FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 9
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-286-981B-9
Query Match 100.0%; Score 889; DB 4; Length 446;
Best Local Similarity 100.0%; Pred. No. 1.2e-65;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IKTDREAEAEAKRRADAKEQCKPKGRAGVPGELATPPDKENDAKSSDSVGEETLPS 60
Db 228 IKTDREAEAEAKRRADAKEQCKPKGRAGVPGELATPPDKENDAKSSDSVGEETLPS 287
QY 61 PSIAKEKVAEAEKVEAKKKAEDQCKEDRRNYPNTNYKTLELEIAESDVVKAELEL 120
Db 268 PSIAKEKVAEAEKVEAKKKAEDQCKEDRRNYPNTNYKTLELEIAESDVVKAELEL 347
QY 121 VEEAKKEPRNEKVKQAFAVESSKKAETRLKIKITDRKKAEEAKRKAEDVKVEKPA 180
Db 348 VEEAKKEPRNEKVKQAFAVESSKKAETRLKIKITDRKKAEEAKRKAEDVKVEKPA 407
RESULT 2
US-08-961-083-38
; Sequence 38, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue

```

CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 453 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-38

Query Match 100.0%; Score 889; DB 3; Length 453;
Best Local Similarity 100.0%; Pred. No. 1.3e-65;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 IKTDREAEAEERADAKGKPKRAGVPGELATPDKKENDAKSSDSVGEETLPS 60
25 IKTDREAEAEERADAKGKPKRAGVPGELATPDKKENDAKSSDSVGEETLPS 84
61 PSIKPEKVAEAKKVEAKKADQKEDRRNPNTYKTLLEIAESDVEVKAELEL 120
85 PSIKPEKVAEAKKVEAKKADQKEDRRNPNTYKTLLEIAESDVEVKAELEL 144
121 VKEAEKPRNEEKVKAKEVESKKAATRLKIKTRKKAEEAKKAAEDVKKEKPA 180
145 VKEAEKPRNEEKVKAKEVESKKAATRLKIKTRKKAEEAKKAAEDVKKEKPA 204

RESULT 3
US-09-536-784-38
Sequence 38, Application US/09536784
Patent No. 6573082
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/536,784
FILING DATE: 30-Oct-1997
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: OCT-30-1997
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB340P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 38:
SEQUENCE CHARACTERISTICS:
LENGTH: 453 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 38:
US-09-536-784-38

Query Match 100.0%; Score 889; DB 4; Length 453;
Best Local Similarity 100.0%; Pred. No. 1.3e-65;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 IKTDREAEAEERADAKGKPKRAGVPGELATPDKKENDAKSSDSVGEETLPS 60
25 IKTDREAEAEERADAKGKPKRAGVPGELATPDKKENDAKSSDSVGEETLPS 84
61 PSIKPEKVAEAKKVEAKKADQKEDRRNPNTYKTLLEIAESDVEVKAELEL 120
85 PSIKPEKVAEAKKVEAKKADQKEDRRNPNTYKTLLEIAESDVEVKAELEL 144
121 VKEAEKPRNEEKVKAKEVESKKAATRLKIKTRKKAEEAKKAAEDVKKEKPA 180
145 VKEAEKPRNEEKVKAKEVESKKAATRLKIKTRKKAEEAKKAAEDVKKEKPA 204

RESULT 4
US-09-286-981B-6
Sequence 6, Application US/09286981B
Patent No. 6503511
GENERAL INFORMATION:
APPLICANT: Wismann, Theresa M.
APPLICANT: Koenig, Scott
APPLICANT: Johnson, Leslie S
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
FILE REFERENCE: 469201-396
CURRENT APPLICATION NUMBER: US/09/286,981B
CURRENT FILING DATE: 1999-04-06
PRIOR APPLICATION NUMBER: US 60/785,743
PRIOR FILING DATE: 1998-05-15
NUMBER OF SEQ ID NOS: 38
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 6
LENGTH: 446
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-286-981B-6

Query Match 99.4%; Score 884; DB 4; Length 446;
Best Local Similarity 99.4%; Pred. No. 3.2e-65;
Matches 179; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 IKTDREAEAEERADAKGKPKRAGVPGELATPDKKENDAKSSDSVGEETLPS 60
228 IKTDREAEAEERADAKGKPKRAGVPGELATPDKKENDAKSSDSVGEETLPS 287
61 PSIKPEKVAEAKKVEAKKADQKEDRRNPNTYKTLLEIAESDVEVKAELEL 120
288 PSIKPEKVAEAKKVEAKKADQKEDRRNPNTYKTLLEIAESDVEVKAELEL 347
121 VKEAEKPRNEEKVKAKEVESKKAATRLKIKTRKKAEEAKKAAEDVKKEKPA 180

Db 348 VKEEAKEPRNEKVKQAKAEVSKKAEATRLKIKTRKKAEBAKKAEDKVEKPA 407

RESULT 5

US-09-286-981B-16

Sequence 16, Application US/09286981B

Patent No. 6503511

GENERAL INFORMATION:

APPLICANT: Wizemann, Theresa M.

APPLICANT: Koenig, Scott

APPLICANT: Johnson, Leslie S

TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines

FILE REFERENCE: 469201-396

CURRENT APPLICATION NUMBER: US/09/286,981B

CURRENT FILING DATE: 1999-04-06

PRIOR APPLICATION NUMBER: US 60/085,743

PRIOR FILING DATE: 1998-05-15

NUMBER OF SEQ ID NOS: 38

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 16

LENGTH: 414

TYPE: PRT

ORGANISM: Streptococcus pneumoniae

US-09-286-981B-16

Query Match 90.4%; Score 803.5; DB 4; Length 414;

Best Local Similarity 86.8%; Pred. No. 1.2e-58;

Matches 164; Conservative 12; Mismatches 4; Indels 9; Gaps 1;

QY 1 IKTDREAEBAEKRRADAK-----EKGKPKGRAGVGPGLATPPDKENDAKSSDS 51

Db 189 IKTRDKKAEBAEKRRADAKLKEANVATSDGKPKGRAGVGPGLATPPDKENDAKSSDS 248

QY 52 SVGEETLPSPSLKPEKVAEAEKVEBAKKAADQKEDRRNYPNTYKTLLEIAESDV 111

Db 249 SVGEETLPSSSLKSGKVAEAEKVEBAKKAADQKEDRRNYPNTYKTLLEIAESDV 308

QY 112 EVKKALELVKEBAEPNEEKVKQAKAEVSKKAEATRLKIKTRKKAEBAKKA 171

Db 309 KYKAELELVKEBAEPDEBKIKQAKAVESKKAATRLIKTRKKAEBAKKA 368

QY 172 EDKVEKPA 180

Db 369 EDKVEKPA 377

QY 172 EDKVEKPA 180

Db 369 EDKVEKPA 377

RESULT 6

US-08-847-065-25

Sequence 25, Application US/08847065

Patent No. 6245335

GENERAL INFORMATION:

APPLICANT: Masure, H. Robert

APPLICANT: Rosenow, Carsten I.

APPLICANT: Tuomaneen, Elaine

APPLICANT: Wizemann, Theresa M.

TITLE OF INVENTION: CHOLINE BINDING PROTEINS FOR

TITLE OF INVENTION: ANTI-PNEUMOCOCCAL VACCINES

NUMBER OF SEQUENCES: 25

CORRESPONDENCE ADDRESS:

ADDRESSEE: David A. Jackson, Esq.

STREET: 411 Hackensack Ave, Continental Plaza, 4th

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/847,065

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-158

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-487-5800

TELEFAX: 201-343-1684

INFORMATION FOR SEQ ID NO: 25:

SEQUENCE CHARACTERISTICS:

LENGTH: 631 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

FRAGMENT TYPE: N-terminal

US-08-847-065-25

Query Match 90.4%; Score 803.5; DB 3; Length 631;

Best Local Similarity 86.8%; Pred. No. 2e-58;

Matches 164; Conservative 12; Mismatches 4; Indels 9; Gaps 1;

QY 1 IKTDREAEBAEKRRADAK-----EKGKPKGRAGVGPGLATPPDKENDAKSSDS 51

Db 155 IKTRDKKAEBAEKRRADAKLKEANVATSDGKPKGRAGVGPGLATPPDKENDAKSSDS 214

QY 52 SVGEETLPSPSLKPEKVAEAEKVEBAKKAADQKEDRRNYPNTYKTLLEIAESDV 111

Db 215 SVGEETLPSSSLKSGKVAEAEKVEBAKKAADQKEDRRNYPNTYKTLLEIAESDV 274

QY 112 EVKKALELVKEBAEPNEEKVKQAKAEVSKKAEATRLKIKTRKKAEBAKKA 171

Db 275 KYKAELELVKEBAEPDEBKIKQAKAVESKKAATRLIKTRKKAEBAKKA 334

QY 172 EDKVEKPA 180

Db 335 EDKVEKPA 343

QY 172 EDKVEKPA 180

Db 335 EDKVEKPA 343

RESULT 7

US-09-308-022-6

Sequence 6, Application US/09308022

Patent No. 6291654

GENERAL INFORMATION:

APPLICANT: REGENTS OF THE UNIVERSITY OF MINNESOTA, ET AL.

TITLE OF INVENTION: C3 BINDING PROTEIN OF STREPTOCOCCUS

TITLE OF INVENTION: PNEUMONIAE

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: METTING, RAASCH & GEBHARDT, P.A.

STREET: 119 No. 6291654th Fourth Street, Suite 203

CITY: Minneapolis

STATE: Minnesota

COUNTRY: USA

ZIP: 55401

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/308,022

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US97/20586

FILING DATE: 12-NOV-1997

APPLICATION NUMBER: 60/029,444

FILING DATE: 12-NOV-1996

APPLICATION NUMBER: 60/038,086

FILING DATE: 18-FEB-1997

APPLICATION NUMBER: 60/059,368
FILING DATE: 19-SEP-1997
APPLICATION NUMBER: 60/062,473
FILING DATE: 16-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: MUETING, Ann M.
REGISTRATION NUMBER: 33,977
REFERENCE/DOCKET NUMBER: 110.00430101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-305-1217
TELEFAX: 612-305-1228
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 564 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
308-022-6

Query Match 90.2%; Score 801.5; DB 3; Length 564;
Best Local Similarity 86.2%; Pred. No. 2.5e-58;
Matches 163; Conservative 13; Mismatches 4; Indels 9; Gaps 1;

QY 1 IKTDREAEAEARRADAK-----EKGKPKGRAGVPGELATPDKKENDAKSSDS 51
DB 226 IKDRKKAEEAKKADAKLEAVATSDGPKGRAGVPGELATPDKKENDAKSSDS 285
QY 52 SVGEETLPSPSLKPEKKVAEAEKKAEADOKEDRRNPPTNTYKTLLELEIAESDV 111
DB 286 SVGEETLPSSSLKSGKVAEAEKKAEKKAKQKEDRRNPPTNTYKTLLEIAESDV 345
QY 112 EVKKAELIYKEAEKPRNEEKVQAAYVESKKAETRLKIKTDKKAEEBAKRAAE 171
DB 346 KVEAELEIYKEAEKPRDEKIKQAKAVESKKAETRLKIKTDKKAEEBAKRAAE 405
QY 172 EDKVKKEKA 180
DB 406 EDKVKKEKA 414

RESULT 8
US-09-286-981B-18
Sequence 18, Application US/09286981B
Patent No. 6503511
GENERAL INFORMATION:
APPLICANT: Wizemann, Theresa M.
APPLICANT: Koenig, Scott
APPLICANT: Johnson, Leslie S.

TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
FILE REFERENCE: 469201-396
CURRENT APPLICATION NUMBER: US/09/286,981B
CURRENT FILING DATE: 1999-04-06
PRIOR APPLICATION NUMBER: US 60/085,743
PRIOR FILING DATE: 1998-05-15
NUMBER OF SEQ ID NOS: 38
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 18
LENGTH: 406
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-286-981B-18

Query Match 89.1%; Score 792; DB 4; Length 406;
Best Local Similarity 90.2%; Pred. No. 1e-57;
Matches 166; Conservative 3; Mismatches 11; Indels 4; Gaps 1;

QY 1 IKTDREAEAEARRADAKQ---KPKGRAGVPGELATPDKKENDAKSSDSVGE 56
DB 188 IKTDREAEAEARRADAKQDSSSKRKRVRGDIQGEQATPDKKENDAKSSDSVGE 247
QY 57 TLSPSLKPEKKVAEAEKKAEKKAEADOKEDRRNPPTNTYKTLLELEIAESDV 116
DB 116 TLSPSLKPEKKVAEAEKKAEKKAEADOKEDRRNPPTNTYKTLLELEIAESDV 116

DB 248 TLSPSLKPEKKVAEAEKKAEADOKEDRRNPPTNTYKTLLELEIAESDV 307
QY 117 ELEIYKEAEKPRNEEKVQAAYVESKKAETRLKIKTDKKAEEBAKRAAE 176
DB 308 ELEIYKEAEKPRNEEKVQAAYVESKKAETRLKIKTDKKAEEBAKRAAE 367
QY 177 EKPA 180
DB 368 EKPA 371

RESULT 9
US-09-286-981B-11
Sequence 11, Application US/09286981B
Patent No. 6503511
GENERAL INFORMATION:
APPLICANT: Wizemann, Theresa M.
APPLICANT: Koenig, Scott
APPLICANT: Johnson, Leslie S.

TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
FILE REFERENCE: 469201-396
CURRENT APPLICATION NUMBER: US/09/286,981B
CURRENT FILING DATE: 1999-04-06
PRIOR APPLICATION NUMBER: US 60/085,743
PRIOR FILING DATE: 1998-05-15
NUMBER OF SEQ ID NOS: 38
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 11
LENGTH: 425
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-286-981B-11

Query Match 89.1%; Score 792; DB 4; Length 425;
Best Local Similarity 85.4%; Pred. No. 1.1e-57;
Matches 164; Conservative 8; Mismatches 8; Indels 12; Gaps 1;

QY 1 IKTDREAEAEARRADAK-----EKGKPKGRAGVPGELATPDKKENDAKS 48
DB 188 IKDRKKAEEAKKADAKLEAVATSDGPKGRAGVPGELATPDKKENDAKS 247
QY 49 SDSSVGEETLPSPSLKPEKKVAEAEKKAEADOKEDRRNPPTNTYKTLLEIAE 108
DB 248 SDSSVGEETLPSPSLKPEKKVAEAEKKAEADOKEDRRNPPTNTYKTLLEIAE 307
QY 109 SDVEVKAELIYKEAEKPRNEEKVQAAYVESKKAETRLKIKTDKKAEEBAKRX 168
DB 308 SDVKAESLEIYKEAEKPRNEEKVQAAYVESKKAETRLKIKTDKKAEEBAKRX 367
QY 169 AAEDKVKKEKA 180
DB 368 AAEDKVKKEKA 379

RESULT 10
US-09-286-981B-10
Sequence 10, Application US/09286981B
Patent No. 6503511
GENERAL INFORMATION:
APPLICANT: Wizemann, Theresa M.
APPLICANT: Koenig, Scott
APPLICANT: Johnson, Leslie S.
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
FILE REFERENCE: 469201-396
CURRENT APPLICATION NUMBER: US/09/286,981B
CURRENT FILING DATE: 1999-04-06
PRIOR APPLICATION NUMBER: US 60/085,743
PRIOR FILING DATE: 1998-05-15
NUMBER OF SEQ ID NOS: 38
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 10
LENGTH: 414
TYPE: PRT

Page 5

57
SHEET

UKAE 117
1111
UKAF 210

KVKE	177
QVKE	370

scines

3; 15

SDS 248
SDV 117

SDV 308
KAA 170

KAA 368

Db 369 EEDKVEKPA 378

RESULT 14
US-09-286-981B-12
Sequence 12, Application US/09286981B
Patent No. 6503511
GENERAL INFORMATION:
APPLICANT: Witzmann, Theresa M.
APPLICANT: Koenig, Scott
APPLICANT: Johnson, Leslie S
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
FILE REFERENCE: 469201-396
CURRENT APPLICATION NUMBER: US/09/286,981B
CURRENT FILING DATE: 1999-04-06
PRIOR APPLICATION NUMBER: US 60/085,743
PRIOR FILING DATE: 1998-05-15
NUMBER OF SEQ ID NOS: 38
SOFTWARE: PatentIn Ver. 2.1
ID NO 12
LENGTH: 426
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-286-981B-12

Query Match 86.7%; Score 770.5; DB 4; Length 426;
Best Local Similarity 87.4%; Pred. No. 6.4e-56;
Matches 166; Conservative 4; Mismatches 9; Indels 11; Gaps 3;

QY 1 IKTDREAEAEAEARRADAK-----EKGKPKRAKRGVPGELATPDKKENDAKSSDS 51
Db 190 IKTDREKA-EAARRADAKLQEANVATSEODKSKRAKREVLGELATPDKKENDAKSSDS 248
QY 52 SVGEETLPSPSLKPEKKVAEAEKKVBAEDQKEDRRNYPNTYKTLLEIAESDV 111
Db 249 SVGEETLPSPSLKPEKKVAEAEKKVBAEDQKEDRRNYPNTYKTLLEIAESDV 308
QY 112 EVKKAELVKEEAKPRNEEKVKQAKAVESKKAATRLKIKTDREKA-EAEARRKAA 170
Db 309 EVKKAELVKEEAKPRNEEKVKQAKAVESKKAATRLKIKTDREKA-EAEARRKAA 368
QY 171 EEDKVEKPA 180
Db 369 EEDKVEKPA 378

SEQ ID NO 13
US-09-286-981B-13
Sequence 13, Application US/09286981B
Patent No. 6503511
GENERAL INFORMATION:
APPLICANT: Witzmann, Theresa M.
APPLICANT: Koenig, Scott
APPLICANT: Johnson, Leslie S
TITLE OF INVENTION: Derivatives of Choline Binding Proteins for Vaccines
FILE REFERENCE: 469201-396
CURRENT APPLICATION NUMBER: US/09/286,981B
CURRENT FILING DATE: 1999-04-06
PRIOR APPLICATION NUMBER: US 60/085,743
PRIOR FILING DATE: 1998-05-15
NUMBER OF SEQ ID NOS: 38
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 13
LENGTH: 425
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-286-981B-13

Query Match 86.6%; Score 769.5; DB 4; Length 425;
Best Local Similarity 87.4%; Pred. No. 7.7e-56;
Matches 166; Conservative 4; Mismatches 9; Indels 11; Gaps 3;

QY 1 IKTDREAEAEAEARRADAK-----EKGKPKRAKRGVPGELATPDKKENDAKSSDS 51

Db 190 IKTDREKA-EAARRADAKLQEANVATSEODKSKRAKREVLGELATPDKKENDAKSSDS 248
QY 52 SVGEETLPSPSLKPEKKVAEAEKKVBAEDQKEDRRNYPNTYKTLLEIAESDV 111
Db 249 SVGEETLPSPSLKPEKKVAEAEKKVBAEDQKEDRRNYPNTYKTLLEIAESDV 308
QY 112 EVKKAELVKEEAKPRNEEKVKQAKAVESKKAATRLKIKTDREKA-EAEARRKAA 170
Db 309 EVKKAELVKEEAKPRNEEKVKQAKAVESKKAATRLKIKTDREKA-EAEARRKAA 368
QY 171 EEDKVEKPA 180
Db 369 EEDKVEKPA 378

Search completed: November 21, 2003, 13:35:11
Job time: 8.55427 secs